

**UNIVERSIDAD AUTONOMA DE MADRID**

**ESCUELA POLITÉCNICA SUPERIOR**



**TRABAJO FIN DE MÁSTER**

# **Resolución de una alerta biológica causada por un patógeno de transmisión alimentaria mediante análisis genómico**

**Máster Universitario en Bioinformática y Biología  
Computacional**

**Autor: González Bodí, Sara**

**Tutor: Cuesta de la Plaza, Isabel y Monzón Fernández, Sara**

**Ponente: Carrillo de Santa Pau, Enrique**

**Dpto.: Unidad de Bioinformática - Instituto de Salud Carlos III**

**Septiembre, 2020**



## **Agradecimientos**

A mis directoras del Trabajo Fin de Máster, Isabel Cuesta y Sara Monzón por su gran apoyo, dedicación y valiosa transferencia de conocimientos.

A mis compañeros de la Unidad de Bioinformática del ISCIII, Sarai Varona y Luis Chapado, por toda su ayuda durante esta etapa de aprendizaje.

A mis compañeros del departamento de Microbiología del Hospital 12 de Octubre, Esther Viedma y Raúl Recio, por todo lo que he aprendido con ellos estos meses y que también he podido aplicar en este trabajo.

A mi familia por estar siempre a mi lado.



## Resumen

---

Las técnicas de secuenciación de alto rendimiento (HTS) han supuesto una revolución científica y médica en muchos campos de las ciencias de la vida. Una de sus aplicaciones más relevantes se centra en el área de la Seguridad Alimentaria y Salud Pública, en el que la prevención, control, gestión y monitorización de los agentes infecciosos productores de las intoxicaciones alimentarias son considerados aspectos fundamentales.

El objetivo principal de este trabajo ha sido mostrar como la implementación de la HTS en la caracterización de una alerta biológica provocada por un brote de Brucelosis puede ser un método altamente discriminatorio de tipificación en tiempo real de los principales patógenos aislados de los enfermos de una toxiinfección alimentaria y de las posibles fuentes de infección.

Para ello se utilizaron los datos de un ejercicio de intercomparación entre laboratorios organizado por la Oficina de Asuntos para el Desarme de las Naciones Unidas (UNODA). Se analizaron 260 genomas completos de *Brucella spp.* y 14 muestras metagenómicas de aislados alimentarios. Para la resolución del ejercicio y con ello la resolución del brote, se diseñó un pipeline de análisis que incluyó, análisis de calidad de los datos, clasificación taxonómica, ensamblado de genomas, análisis de polimorfismos y filogenético, y caracterización del resistoma, viruloma y plasmidoma de las muestras recibidas para analizar su inclusión en el brote de toxiinfección alimentaria.

De acuerdo a los resultados obtenidos, la especie causante del brote se confirma como *B. melitensis*, y de acuerdo a la relación filogenética se observó cómo la mayor parte de los aislados clínicos, clasificados como *B. melitensis*, se clasifican junto con un aislado alimentario y una muestra metagenómica de la misma especie agrupándose en una rama diferenciada, y conformando un grupo monofilético determinado por una única cepa. Estos resultados permitieron determinar la especie vinculada al brote y discriminar con una alta capacidad de resolución aquellos aislados clínicos pertenecientes al brote de *B. melitensis*. Asimismo, se confirmó su dispersión a otras muestras alimentarias y se determinó el origen alimentario del brote.

En conclusión, este trabajo ha permitido validar y poner a punto un pipeline de trabajo que está disponible públicamente en la plataforma GitHub de la unidad de bioinformática del ISCIII y GitHub del estudiante de este trabajo que podrá ser utilizado para labores de vigilancia de salud pública en estudios de identificación, seguimiento, evolución y trazabilidad de brotes infecciosos causados por patógenos.

**Palabras clave:** investigación de brotes de origen alimentario; secuenciación del genoma completo; metagenómica; filogenia



## Abstract

---

High-throughput sequencing (HTS) techniques represent a scientific and medical revolution in many fields of the life sciences. One of the most relevant applications is in the area of Food Safety and Public Health, where the prevention, control, management and monitoring of foodborne pathogens are considered crucial aspects.

The main objective of this work has been to demonstrate that the implementation of HTS techniques in the resolution of biological alert produced by an outbreak of Brucellosis can be used as an accurate, reliable and highly discriminative subtyping method for epidemiological characterization and outbreak strain investigation.

For this propose, datasets of an interlaboratory comparison exercise organized by United Nations Office for Disarmament Affairs (UNODA) were used. A total of 260 single genomes of *Brucella spp.* and 14 metagenomic samples considered as whole community sequence data has been analyzed. In order to solve the exercise and therefore the outbreak, a work pipeline was designed including data quality analysis, taxonomic classification, genome assembly, core single-nucleotide polymorphisms and phylogenetic analysis, and characterization of the resistome, viruloma and plasmidome of the samples studied. This allowed to identify which samples are related to the outbreak and the source of that.

According to the results obtained in phylogenetic cluster analysis, it was observed that most of the clinical isolates, classified as *B. melitensis* specie, together with a food isolate and a metagenomic sample were grouped into a separated branch forming a monophyletic group determined by a single strain. These results provided definitive evidence of the specific *Brucella* specie outbreak isolated as well as the number of human cases involved in the *B. melitensis* outbreak. Furthermore, a not human origin genome was also related to human outbreak and food source from metagenomic dataset was confirmed.

In conclusion, this work permitted to create and validate a work pipeline that is open access on both ISCIII bioinformatics unit and student's GitHub platform. This tool aims to be useful for public health surveillance tasks in studies of identification, monitoring, evolution and traceability of infectious outbreaks caused by pathogens.

**Key words:** foodborne outbreak investigation; whole genome sequencing; metagenomics; phylogeny





RESUMEN .....	5
ABSTRACT .....	7
ÍNDICE GENERAL .....	9
ÍNDICE DE FIGURAS .....	11
ÍNDICE DE TABLAS .....	13
ÍNDICE DE TABLAS ANEXOS .....	15
1 . INTRODUCCIÓN .....	17
2. HIPÓTESIS Y OBJETIVOS.....	21
3. METODOLOGÍA Y PLAN DE TRABAJO .....	22
4. MATERIAL Y MÉTODOS .....	23
4. 1 ESCENARIO.....	23
4. 2 DATOS RECIBIDOS PARA EL ANÁLISIS .....	24
4.3 DESARROLLO DEL FLUJOGRAMA DE TRABAJO PROPUESTO PARA EL ANÁLISIS GENÓMICO Y METAGENÓMICO DEL ESTUDIO. ....	25
4.3.1 DATOS CRUDOS Y LISTADO DE MUESTRAS .....	26
4.3.2 CONTROL Y FILTRADO POR CALIDAD.....	26
4.3.3 IDENTIFICACIÓN DE ESPECIES BACTERIANAS - ANÁLISIS TAXONÓMICO .....	28
4.3.4 ENSAMBLADO.....	30
4.3.5 ANÁLISIS FILOGENÉTICO DE LOS AISLAMIENTOS VINCULADOS EPIDEMIOLÓGICAMENTE MEDIANTE ANÁLISIS DE POLIMORFISMOS DE NUCLEÓTIDO ÚNICO (SNPs).....	31
4.3.5 CARACTERIZACIÓN DE LOS AISLADOS Y MUESTRAS METAGENÓMICAS VINCULADAS AL BROTE DE BRUCELOSIS. ....	33
4.3.6 VISUALIZACIÓN DE LOS RESULTADOS.....	34
5. RESULTADOS.....	35
5.1 ANÁLISIS DE CALIDAD DE LAS LECTURAS FILTRADAS (FICHEROS FASTQC) .....	35

<b>5. 2 ANÁLISIS TAXONÓMICO.....</b>	<b>38</b>
<b>5.3 ANÁLISIS DE CALIDAD LOS ENSAMBLADOS DEL COMPONENTE 1 (FICHEROS FASTA).....</b>	<b>40</b>
<b>5.3 ANÁLISIS FILOGENÉTICO .....</b>	<b>42</b>
<b>5.4 CARACTERIZACIÓN DE LOS AISLADOS Y MUESTRAS METAGENÓMICAS VINCULADAS AL BROTE DE BRUCELOSIS44</b>	
<b>5.4.1 TIPIFICACIÓN MULTILOCUS DE SECUENCIAS, MLST (ST). ....</b>	<b>44</b>
<b>5.4.2 GENES DE VIRULENCIA Y RESISTENCIA A ANTIBIÓTICOS.....</b>	<b>45</b>
<b>6. CONCLUSIONES Y PERSPECTIVAS FUTURAS .....</b>	<b>47</b>
<b>8. ANEXOS.....</b>	<b>49</b>

<b>FIGURA 1.</b> ENCUADRE TAXONÓMICO Y PRINCIPALES CARACTERÍSTICAS DE LAS ESPECIES DEL GÉNERO <i>BRUCELLA</i> . ...	<b>18</b>
<b>FIGURA 2.</b> CURVAS EPIDÉMICAS. <b>A.</b> EVOLUCIÓN DE CASOS DE BRUCELOSIS REPORTADOS DURANTE EL 2018. <b>B.</b> EVOLUCIÓN DE CASOS DE BRUCELOSIS REPORTADOS DURANTE EL PERIODO 2010-2013 EN COMPARACIÓN CON 2014.....	<b>23</b>
<b>FIGURA 3.</b> DISTRIBUCIÓN GEOGRÁFICA DE CASOS DE ENFERMEDAD EN 2018 EN EL ÁREA RELEVANTE. ....	<b>24</b>
<b>FIGURA 4.</b> ESQUEMA GENERAL DEL ANÁLISIS BIOINFORMÁTICO. ....	<b>25</b>
<b>FIGURA 5.</b> ESQUEMA DEL PROCESAMIENTO DE LECTURAS EN MODO PAIRED-END. TRIMMOMATIC. ....	<b>27</b>
<b>FIGURA 6.</b> REPRESENTACIÓN DEL ALGORITMO UTILIZADO EN LA HERRAMIENTA KRAKEN. ....	<b>28</b>
<b>FIGURA 6.</b> EJEMPLO DE INFORME DE RESULTADOS DE KRAKEN. SE TRATA DE UN ARCHIVO DELIMITADO POR TABULACIONES, CON UNA LÍNEA POR TAXÓN. LOS CAMPOS DE LA SALIDA, DE IZQUIERDA A DERECHA SON: PORCENTAJE DE LECTURAS DEL CLADO EN ESTE TAXÓN, NÚMERO DE LECTURAS TOTALES DEL CLADO EN ESTE TAXÓN, NÚMERO DE LECTURAS ASIGNADAS ESPECÍFICAMENTE A ESTE TAXÓN, RANGO DE CÓDIGO DE LETRAS QUE INDICAN (U) NO CLASIFICADO, (D) DOMINIO, (K) REINO, (P) FILO / DIVISIÓN, (C) CLASE, (O) ORDEN, (F) FAMILIA, (G) GENERO, O (S) ESPECIE, NCBI ID TAXONÓMICO (BASE DE DATOS TAXONOMY; <a href="https://www.ncbi.nlm.nih.gov/taxonomy">HTTPS://WWW.NCBI.NLM.NIH.GOV/TAXONOMY</a> ), DESCRIPCIÓN. ....	<b>29</b>
<b>FIGURA 7.</b> REPRESENTACIÓN DE GRAFO DE-BRUIJIN. ....	<b>30</b>
<b>FIGURA 8.</b> PHRED SCORE POR BASE.....	<b>35</b>
<b>FIGURA 9.</b> PHRED SCORE POR LECTURAS.....	<b>36</b>
<b>FIGURA 10.</b> CONTENIDO DE GC% POR SECUENCIA .....	<b>36</b>
<b>FIGURA 11.</b> PHRED SCORE POR BASE.....	<b>37</b>
<b>FIGURA 12.</b> PHRED SCORE POR LECTURAS.....	<b>38</b>
<b>FIGURA 13.</b> CONTENIDO DE GC% POR SECUENCIA. ....	<b>38</b>
<b>FIGURA 14.</b> DISTRIBUCIÓN DE ESPECIES DE ACUERDO AL TIPO DE MUESTRA. ....	<b>39</b>
<b>FIGURA 15.</b> EJEMPLO DE RESULTADOS DE KRAKEN VISUALIZADOS EN LA HERRAMIENTA INTERACTIVA KRONA. ....	<b>40</b>
<b>FIGURA 16.</b> ÁRBOL FILOGENÉTICO KSNP DE MÁXIMA VEROSIMILITUD (ML) GENERADO MEDIANTE RAXML A PARTIR DE LA ALINEACIÓN DE LOS SNPs CORE GENERADOS POR KSNP. SE AMPLÍA EL CLADO PERTENECIENTE A LA SECUENCIA TIPO 11, DONDE SE AGRUPAN LA MAYOR PARTE DE LAS MUESTRAS DE AISLADOS HUMANOS DE <i>B. MELITENSIS</i> . ...	<b>42</b>
<b>FIGURA 17.</b> ÁRBOL FILOGENÉTICO SNIPPY RECONSTRUIDO A PARTIR DE UN ALINEAMIENTO DE LOS SNPs CONSERVADO GENERADO EN BASE AL MAPEO DE LECTURAS FILTRADAS FRENTE A LA REFERENCIA Y CORREGIDO POR RECOMBINACIÓN USANDO GUBBINS. ....	<b>43</b>



<b>TABLA 1.</b> CASOS REPORTADOS DE ZOONOSIS EN EU EN EL 2018. ....	<b>17</b>
<b>TABLA 2.</b> METODOLOGÍA Y PLAN DE TRABAJO. ....	<b>22</b>
<b>TABLA 3.</b> EJEMPLO DE INFORME DE RESULTADOS DE BRACKEN. LOS CAMPOS DE LA SALIDA, DE IZQUIERDA A DERECHA SON: DESCRIPCIÓN, NCBI ID TAXONÓMICO, RANGO DE CÓDIGO DE LETRAS QUE INDICAN (U) NO CLASIFICADO, (D) DOMINIO, (K) REINO, (P) FILO / DIVISIÓN, (C) CLASE, (O) ORDEN, (F) FAMILIA, (G) GENERO, O (S) ESPECIE, LECTURAS ASIGNADAS POR KRAKEN2, LECTURAS AÑADIDAS, NUEVAS LECTURAS ASIGNADAS, DATO DE ABUNDANCIA. ....	<b>29</b>
<b>TABLA 4.</b> DESCRIPCIÓN DE LOS PARÁMETROS DE MÁS RELEVANTES DE QUASt. ....	<b>31</b>
<b>TABLA 5.</b> GENES DE MANTENIMIENTO UTILIZADOS PARA LA TIPIFICACIÓN POR MLST. ....	<b>34</b>
<b>TABLA 6.</b> DATOS DE CALIDAD DEL COMPONENTE 2 VISUALIZADOS CON MULTIQC. ....	<b>37</b>
<b>TABLA 7.</b> IDENTIFICACIÓN DE LA ESPECIE MAYORITARIA DE LAS MUESTRAS DEL COMPONENTE 2 ....	<b>39</b>
<b>TABLA 8.</b> PARÁMETROS DE CALIDAD DE LOS ENSAMBLADOS DEL COMPONENTE 1 ....	<b>40</b>
<b>TABLA 9.</b> PRINCIPALES PARÁMETROS DE CALIDAD DE UN SUBGRUPO DE MUESTRAS DEL COMPONENTE 1. ....	<b>41</b>
<b>TABLA 10.</b> PARÁMETROS DE CALIDAD DE LOS ENSAMBLADOS DEL COMPONENTE 1 EXCLUYENDO MUESTRAS QUE PRESENTARON UN PEOR ENSAMBLADO. ....	<b>41</b>
<b>TABLA 11.</b> DISTRIBUCIÓN DE STs EN LAS MUESTRAS DEL COMPONENTE 1. ....	<b>44</b>
<b>TABLA 12.</b> DISTRIBUCIÓN Y FRECUENCIAS DE ST EN LOS AISLADOS DE <i>B. MELITENSIS</i> (N= 129) DE ACUERDO AL TIPO DE MUESTRA. ....	<b>45</b>
<b>TABLA 13.</b> GENES DE VIRULENCIA EN LAS CEPAS CAUSANTES DEL BROTE DE BRUCELOSIS. ....	<b>45</b>
<b>TABLA 14.</b> GENES DE RESISTENCIA A ANTIBIÓTICOS EN LAS CEPAS CAUSANTES DEL BROTE DE BRUCELOSIS. ....	<b>46</b>



---

<b>TABLA ANEXO 1.</b> PARÁMETROS DE CALIDAD PRINCIPALES DE LOS FICHEROS FASTQ DEL COMPONENTE 1 Y 2 (PARSE_FASTQC.PY). .....	<b>49</b>
<b>TABLA ANEXO 2.</b> PARÁMETROS DE CALIDAD PRINCIPALES DE LOS FICHEROS FASTA DEL COMPONENTE2 (PARSE_QUAST.PY). .....	<b>55</b>
<b>TABLA ANEXO 3.</b> RESULTADOS DEL ANÁLISIS TAXONÓMICO DE KRAKEN Y BRACKEN PARA MUESTRAS DEL COMPONENTE 1. EN LA TABLA APARECEN RESUMIDOS LOS PARÁMETROS DE HIT CON MAYOR NÚMERO DE LECTURAS ASIGNADAS (PARSE_KRAKEN.PY). .....	<b>61</b>
<b>TABLA ANEXO 4.</b> RESULTADOS DEL ANÁLISIS TAXONÓMICO DE KRAKEN Y BRACKEN PARA MUESTRAS DEL COMPONENTE 2. EN LA TABLA APARECEN RESUMIDOS LOS PARÁMETROS DEL BEST Y SECOND HIT CON MAYOR NÚMERO DE LECTURAS ASIGNADAS (PARSE_KRAKEN.PY).....	<b>67</b>
<b>TABLA ANEXO 5.</b> RESULTADOS DEL ANÁLISIS DE TIPIFICACIÓN MULTILOCUS DE SECUENCIAS. ....	<b>68</b>





# 1 . Introducción

Las técnicas de secuenciación de alto rendimiento (HTS) inicialmente denominadas NGS (Next Generation Sequencing), han supuesto una revolución científica y médica en muchos campos de las ciencias de la vida. Una de sus aplicaciones más relevantes es en el área de la Seguridad Alimentaria y Salud Pública, en el que la prevención, control, gestión y monitorización de los agentes infecciosos productores de las intoxicaciones alimentarias son considerados aspectos fundamentales<sup>1</sup>. La HTS agiliza considerablemente la caracterización genómica de aislados y comunidades microbianas de interés agroalimentario, facilitando así la identificación de la fuente de contaminación y la evaluación de los riesgos asociados a los agentes contaminantes.<sup>2</sup>

En Europa según un informe publicado por la Agencia de Seguridad Alimentaria Europea (EFSA) los principales patógenos causantes de toxiinfecciones alimentarias (TIAs) en el año 2018 han sido *Campylobacter* spp., *Salmonella* spp, y *Escherichia coli* productora de shiga toxina (STEC), convirtiéndolos en patógenos de especial vigilancia y siendo una prioridad no sólo por su impacto en Salud Pública sino también por el impacto económico que produce en los países a los que afectan las zoonosis.<sup>3</sup>

**Tabla 1.** Casos reportados de zoonosis en EU en el 2018.

Disease	Number of confirmed human cases	Hospitalisation				Deaths			
		Status available (%)	Number of reporting MS <sup>(b)</sup>	Reported hospitalised cases	Proportion hospitalised (%)	Outcome available (%)	Number of reporting MS <sup>(b)</sup>	Reported deaths	Case fatality (%)
<b>Campylobacteriosis</b>	246,571	27.7	18	20,948	30.6	72.7	16	60	0.03
<b>Salmonellosis</b>	91,857	43.2	15	16,556	41.7	67.0	17	119	0.19
<b>STEC infections</b>	8,161	37.3	18	1,151	37.8	60.4	20	11	0.22
<b>Yersiniosis</b>	6,699	26.4	14	519	29.3	56.8	15	3	0.08
<b>Listeriosis</b>	2,549	42.4	17	1,049	97.0	57.6	19	229	15.6
<b>West Nile fever<sup>(a)</sup></b>	1,548	44.7	10	634	91.6	84.6	11	137	10.5
<b>Echinococcosis</b>	793	29.6	13	109	46.4	35.1	14	3	1.10
<b>Q fever</b>	789	NA <sup>(c)</sup>	NA	NA	NA	58.7	12	8	1.70
<b>Brucellosis</b>	358	44.4	9	159	71.1	29.9	10	1	0.93
<b>Tularaemia</b>	300	32.0	11	69	71.9	47.7	11	0	0.0
<b>Trichinellosis</b>	66	21.2	6	9	64.3	27.3	6	0	0.0
<b>Rabies</b>	1	NA <sup>(c)</sup>	NA	NA	NA	100.0	1	1	100.0

MS: Member State.

(a): Instead of confirmed human cases, the total number of human cases were included.

(b): Not all countries observed cases for all diseases.

(c): NA: Not applicable as the information is not collected for this disease.

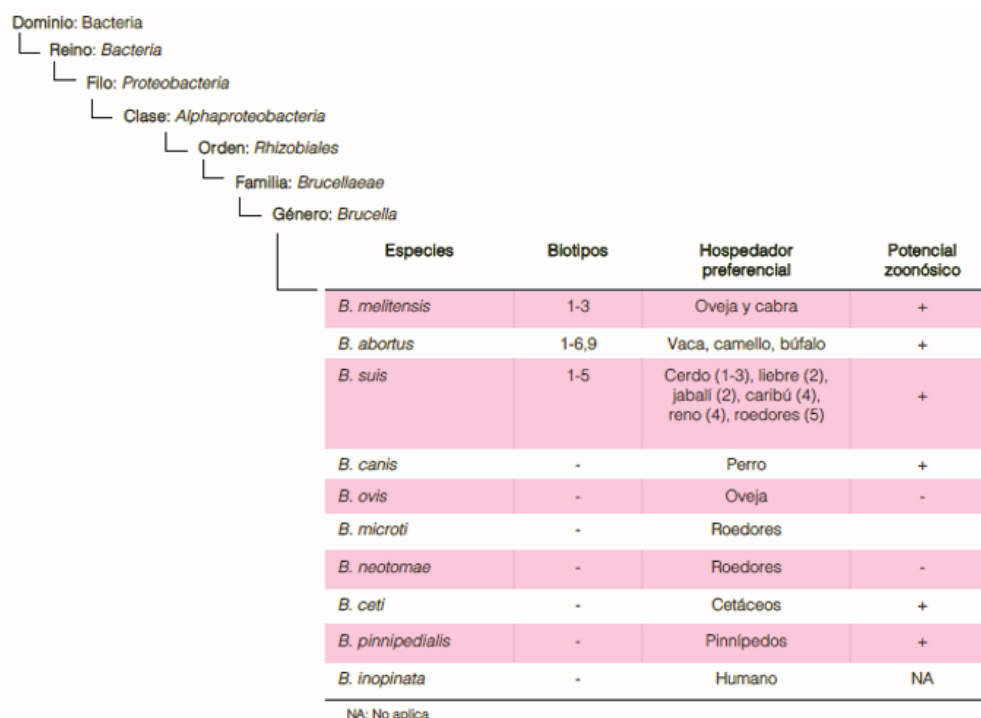
Por otro lado, a nivel global, cabe destacar la importancia de la brucelosis ya que es considerada una de las enfermedades zoonóticas más frecuentes en la mayor parte del mundo que afecta principalmente a países en desarrollo siendo un gran problema de salud pública<sup>4</sup>. Se conocen diez especies de *Brucella*, algunas de ellas con distintos biotipos, cada una de las cuales se diferencia en la especificidad del huésped (Figura 1).

<sup>1</sup> Collineau, L., Boerlin, P., Carson, C. A., Chapman, B., Fazil, A., Hetman, B., McEwen, S. A., Parmley, E. J., Reid-Smith, R. J., Taboada, E. N., & Smith, B. A. (2019). Integrating Whole-Genome Sequencing Data Into Quantitative Risk Assessment of Foodborne Antimicrobial Resistance: A Review of Opportunities and Challenges. *Frontiers in microbiology*, 10, 1107.

<sup>2</sup> Vohra, P., Bugarel, M., Turner, F., Loneragan, G. H., Hope, J. C., Hopkins, J., & Stevens, M. P. (2018). Quantifying the Survival of Multiple *Salmonella enterica* Serovars In Vivo via Massively Parallel Whole-Genome Sequencing To Predict Zoonotic Risk. *Applied and environmental microbiology*, 84(4), e02262-17.

<sup>3</sup> EFSA and ECDC (European Food Safety Authority and European Centre for Disease Prevention and Control), 2019. The European Union One Health 2018 Zoonoses Report. *EFSA Journal* 2019;17(12):5926,276pp.

<sup>4</sup> Pappas G, Papadimitriou P, Akritidis N, Christou L, Tsianos EV. The new global map of human brucellosis. *Lancet Infect Dis*. 2006;6(2):91-99. doi:10.1016/S1473-3099(06)70382-6.



**Figura 1.** Encuadre taxonómico y principales características de las especies del género *Brucella*.<sup>5</sup>

En los animales esta enfermedad da lugar a abortos e infertilidad principalmente, mientras que en las personas la enfermedad se manifiesta como un síndrome febril que puede progresar a una fase crónica caracterizada por la aparición de severas complicaciones como endocarditis, artralgia y epididimitis entre otras. Tres especies de *Brucella* son las que afectan predominantemente al ser humano: *B. melitensis*, *B. abortus* y *B. suis*. De las infecciones causadas por estas tres especies, las debidas a *B. melitensis* son las observadas con mayor frecuencia en los seres humanos, y además son las más graves<sup>6</sup>.

Una de las claves para el reconocimiento y control de estos brotes radica en poder disponer de métodos altamente discriminatorios de tipificación en tiempo real de los principales patógenos aislados de los enfermos de una toxiinfección alimentaria y de las posibles fuentes de infección<sup>7</sup>.

Tradicionalmente la electroforesis en gel de campo pulsado (PFGE) se ha considerado el método standard de tipificación molecular bacteriana<sup>8</sup>. Sin embargo, la capacidad de discriminación entre cepas de esta técnica es muy limitada si se compara con el análisis del genoma basado en SNPs o en alelos<sup>9</sup>. En este contexto, la secuenciación del genoma

<sup>5</sup> Godfroid J, Al Dahouk S, Pappas G, et al. A "One Health" surveillance and control of brucellosis in developing countries: moving away from improvisation. *Comp Immunol Microbiol Infect Dis*. 2013;36(3):241-248.

<sup>6</sup> Seleem MN, Boyle SM, Sriranganathan N. Brucellosis: a re-emerging zoonosis. *Vet Microbiol*. 2010;140(3-4):392-398.

<sup>7</sup> European Centre for Disease Prevention and Control. ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations –2019–2021. Stockholm: ECDC; 2019.

<sup>8</sup> Neoh HM, Tan XE, Sapri HF, Tan TL. Pulsed-field gel electrophoresis (PFGE): A review of the "gold standard" for bacteria typing and current alternatives. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*. 2019; 74:103935.

<sup>9</sup> Quijada NM, Rodríguez-Lázaro D, Eiros JM, Hernández M. TORMES: an automated pipeline for whole bacterial genome analysis. *Bioinformatics*. 2019;35(21):4207-4212.

(WGS) ha dinamizado el campo de la epidemiología molecular siendo significativamente útil para investigar brotes de enfermedades transmitidas por alimentos ya que frecuentemente las cepas involucradas son de tipo monomórficas o clonales<sup>10</sup>. En la actualidad, la administración de Alimentos y Medicamentos (FDA) y los Centros para el Control y la Prevención de Enfermedades de los Estados Unidos (CCPEEU) utilizan WGS como método para caracterizar alimentos y aislados clínicos de *Salmonella* y *Listeria monocytogenes*. Asimismo, la EFSA publicó recientemente un informe científico recomendando la introducción de esta técnica en las actividades de control alimentario donde promueve su futura implantación en la normativa europea del 2021 para la monitorización de la resistencia a los agentes antimicrobianos (RAM)<sup>11</sup>. Actualmente, WGS es considerada la herramienta más completa (*all-in-one test*) ya que permite recoger toda la información que obtenemos de forma independiente con diferentes métodos tipificación en un único abordaje convirtiéndose en una metodología automatizada, rápida y de bajo costo<sup>12</sup>.

Otra de las aplicaciones importantes que tiene la secuenciación masiva en el campo de la detección de los agentes infecciosos es la metagenómica, o lo que es lo mismo, la caracterización genómica de los diferentes microorganismos que componen una muestra sin pasar por el aislamiento de cada uno de ellos, identificando tanto patógenos conocidos como desconocidos<sup>13</sup>. En este sentido, la Organización Mundial de Sanidad Animal (OIE) estima que el 60% de las enfermedades infecciosas en humanos, y el 75% de las emergentes son de origen zoonótico, y de éstos aproximadamente el 60% son causadas por patógenos desconocidos<sup>14</sup>, por lo que la metagenómica se convierte en una tecnología de gran utilidad en el estudio de brotes provocados por patógenos que se transmiten por alimentos o de origen zoonótico<sup>15</sup>.

Diferentes organismos internacionales responsables de la seguridad alimentaria, de la respuesta ante una alerta biológica y del control de las enfermedades infecciosas en los distintos ámbitos, humano y animal, y atendiendo al concepto de One Health, están haciendo esfuerzos por incorporar la secuenciación masiva como metodología de respuesta. En ese esfuerzo han generado diferentes estrategias de armonización y estandarización de los análisis genómicos. Una de ellas es el planteamiento de ejercicios de simulación de alertas biológicas para comprobar la capacitación de los diferentes países en la respuesta. Así, la Oficina de Asuntos para el Desarme de las Naciones Unidas (UNODA) ha puesto en marcha un ejercicio simulacro de alerta biológica en el que la Unidad de Bioinformática del ISCIII ha participado, con el objetivo de prepararse y poder asistir a los laboratorios del Centro Nacional de Microbiología del ISCIII, responsables de

<sup>10</sup> Deng X, Desai PT, den Bakker HC, et al. Genomic epidemiology of *Salmonella enterica* serotype Enteritidis based on population structure of prevalent lineages. *Emerg Infect Dis.* 2014;20(9):1481-1489.

<sup>11</sup> EFSA (European Food Safety Authority), Aerts, M., Battisti, A., Hendriksen, R., Kempf, I., Teale, C., Tenhagen, B-A., Veldman, K., Wasyl, D., Guerra, B., Liebana, E., Thomas-Lopez D and Belœil, P-A. (2019). Scientific report on the technical specifications on harmonised monitoring of antimicrobial resistance in zoonotic and indicator bacteria from food-producing animals and food. *EFSA Journal* 2019;17(6):5709, 122 pp.

<sup>12</sup> Brown E, Dessai U, McGarry S, Gerner-Smidt P. Use of Whole-Genome Sequencing for Food Safety and Public Health in the United States. *Foodborne Pathog Dis.* 2019;16(7):441-450. doi:10.1089/fpd.2019.2662

<sup>13</sup> Mintzer V, Moran-Gilad J, Simon-Tuval T. Operational models and criteria for incorporating microbial whole genome sequencing in hospital microbiology - A systematic literature review. *Clin Microbiol Infect.* 2019;25(9):1086-1095.

<sup>14</sup> Scallan E, Hoekstra RM, Angulo FJ, et al. Foodborne illness acquired in the United States--major pathogens. *Emerg Infect Dis.* 2011;17(1):7-15.

<sup>15</sup> Kovac J. Precision Food Safety: a Paradigm Shift in Detection and Control of Foodborne Pathogens. *mSystems.* 2019;4(3):e00164-19.

dar respuesta en estas situaciones. El escenario de alerta biológica servirá como material para el desarrollo de la práctica propuesta para este trabajo de fin de máster.

## 2. Hipótesis y Objetivos

---

La implementación de la HTS como técnica de tipificación facilitará la detección e identificación de brotes causados por agentes zoonóticos, con mayor poder de discriminación que las técnicas clásicas hasta ahora utilizadas, permitiendo así una respuesta rápida, que facilitará la identificación de los riesgos emergentes potenciales y permitirá la aplicación de medidas preventivas de vigilancia y control que protejan la Salud Pública y Animal, reduciendo al mínimo el perjuicio económico.

Para probar esta hipótesis se ha planteado resolver un ejercicio del UNODA donde el principal objetivo es caracterizar una alerta biológica producida por un brote de Brucelosis ocurrido en un área geográfica concreta y determinar su origen. Para ello se desarrollará un pipeline de análisis de datos de secuenciación de genomas completos y de metagenómica.

Para ello se establecen los siguientes objetivos parciales:

- **Objetivo 1.** Control de calidad de las secuencias entregadas para la resolución del ejercicio.
- **Objetivo 2.** Relaciones filogenéticas entre los aislados del Componente 1 (Genomas completos) para determinar los aislados pertenecientes al brote.
- **Objetivo 3.** Análisis metagenómico del Componente 2 para determinar el origen del brote.

### 3. Metodología y plan de trabajo

En la tabla 2 se representa la planificación en horas para la realización del trabajo fin de máster, organizada según las tareas a desarrollar. Esta organización se realiza teniendo en cuenta la complejidad y necesidad de tiempo de cada uno de los apartados.

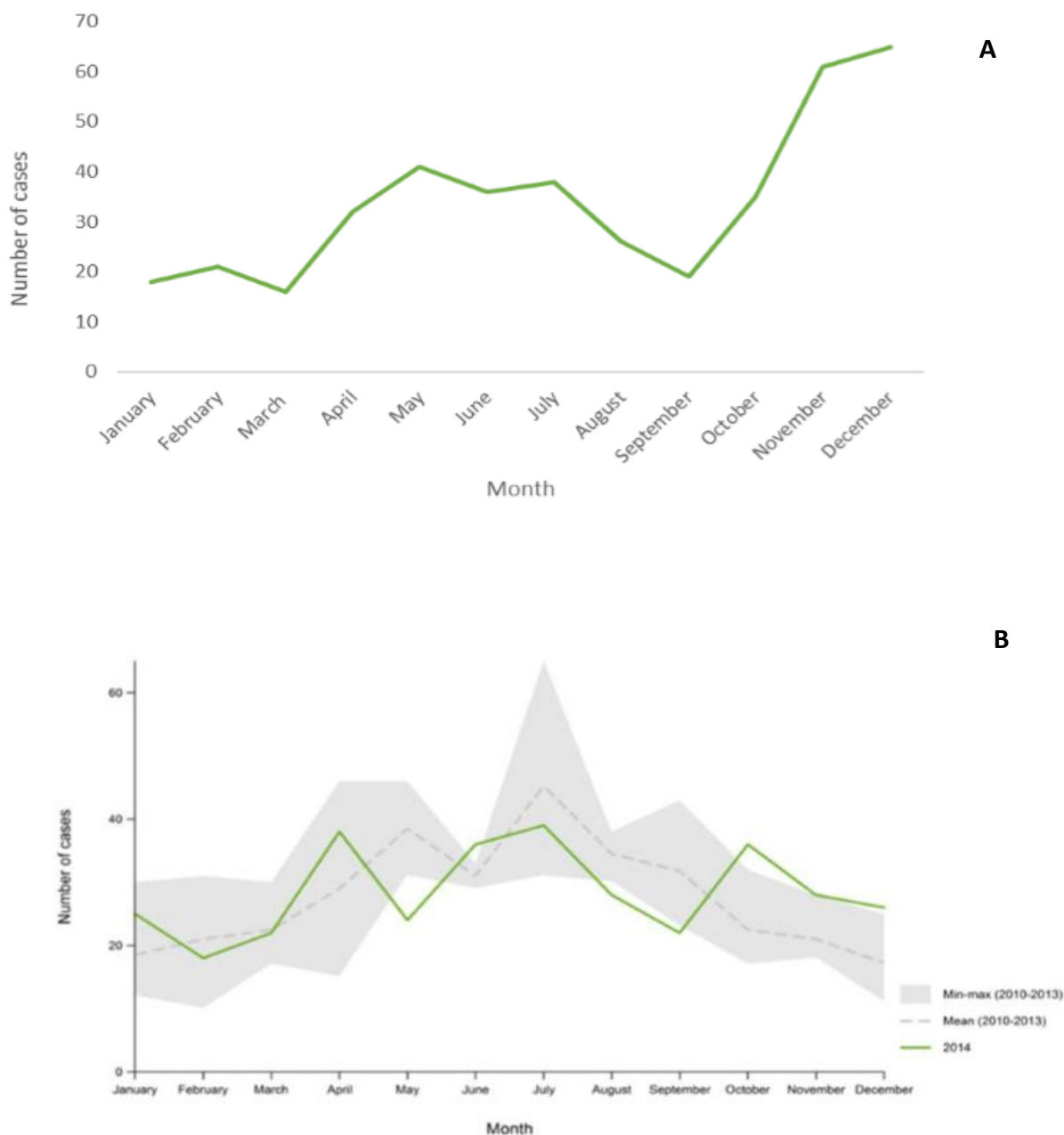
**Tabla 2.** Metodología y plan de trabajo.

Tareas / subtareas	Horas
<b>T1. Estudio del brote propuesto simulado provocado por el género <i>Brucella</i>.</b>	25h (2 semanas)
T1.1 Descarga de genomas disponibles del género <i>Brucella</i> y homólogos	
T1.2 Brotes causados por <i>Brucella</i> - Revisión de bibliografía	
T1.3 Aplicación de HTS en <i>Brucella</i>	
<b>T2. Desarrollo de un pipeline de análisis de datos de WGS</b>	100h (8 semanas)
T2.1 Implementación de: identificación de especie y control de contaminación.	
T2.2 Tipificación basada en SNPs	
<b>T3. Análisis metagenómico enfocado a la identificación de patógenos</b>	100h (8 semanas)
<b>T4. Análisis del brote de simulación</b>	25h (2 semanas)
<b>T5. Escritura de la memoria TFM</b>	40h (4 semanas)
<b>T6. Desarrollo de la presentación y defensa pública del proyecto</b>	10h (2 semanas)
<b>Horas totales</b>	<b>300h (24 semanas)</b>

## 4. Material y Métodos

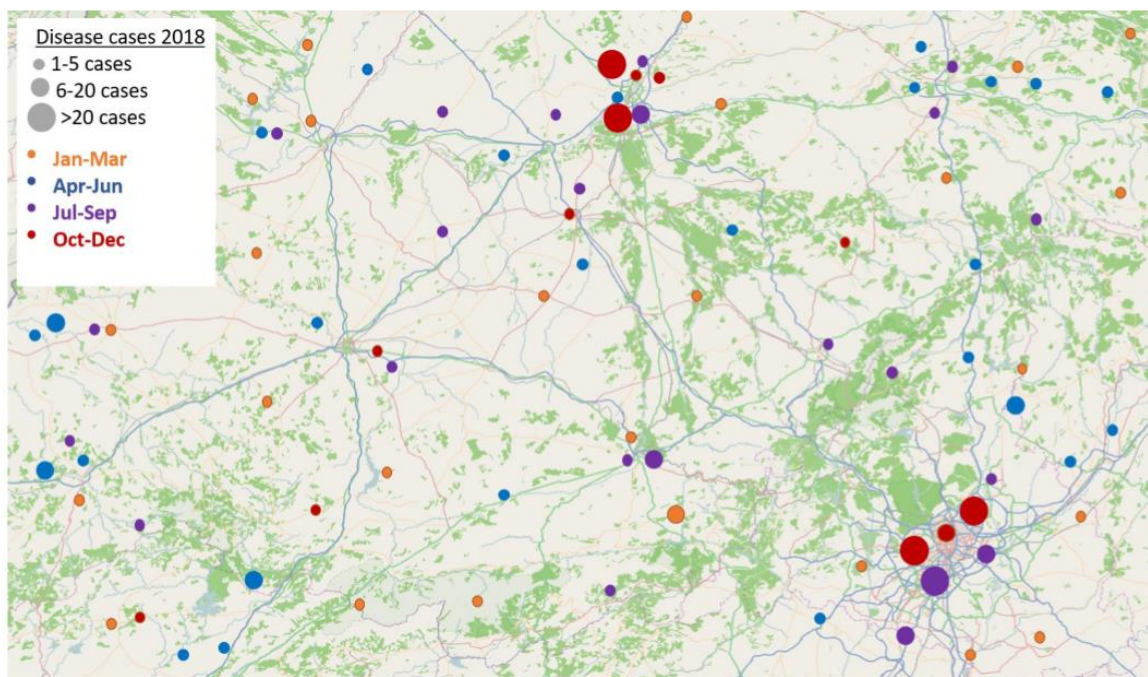
### 4.1 Escenario

En un área geográfica conocida por tener casos endémicos y esporádicos de brucelosis durante el verano, se registró un claro aumento del número de casos de personas enfermas a finales de otoño. La figura 2 muestra la incidencia de casos de Brucelosis en el año 2018, en comparación con años anteriores.



**Figura 2.** Curvas epidémicas. **A.** Evolución de casos de brucelosis reportados durante el 2018. **B.** Evolución de casos de brucelosis reportados durante el periodo 2010-2013 en comparación con 2014.

La mayoría de los pacientes recientemente enfermos pertenecían a dos ciudades concretas vinculadas a eventos deportivos importantes próximos. La ruta de transmisión era desconocida. Las anomalías en el patrón geográfico, el alto número inusual de casos humanos y la prolongada temporada de brotes llevaron a los epidemiólogos a extender la investigación para incluir la detección de brotes con secuenciación del genoma de un número limitado de aislamientos clínicos relacionados epidemiológicamente (Figura 2).



**Figura 3.** Distribución geográfica de casos de enfermedad en 2018 en el área relevante.

#### 4. 2 Datos recibidos para el análisis

Para la realización del estudio de este brote se recibieron dos paquetes de datos identificados como Componente 1 y Componente 2. Las muestras recogidas en ambos componentes se seleccionaron para su análisis basándose en la relación epidemiológica que presentaban en los estudios previos.

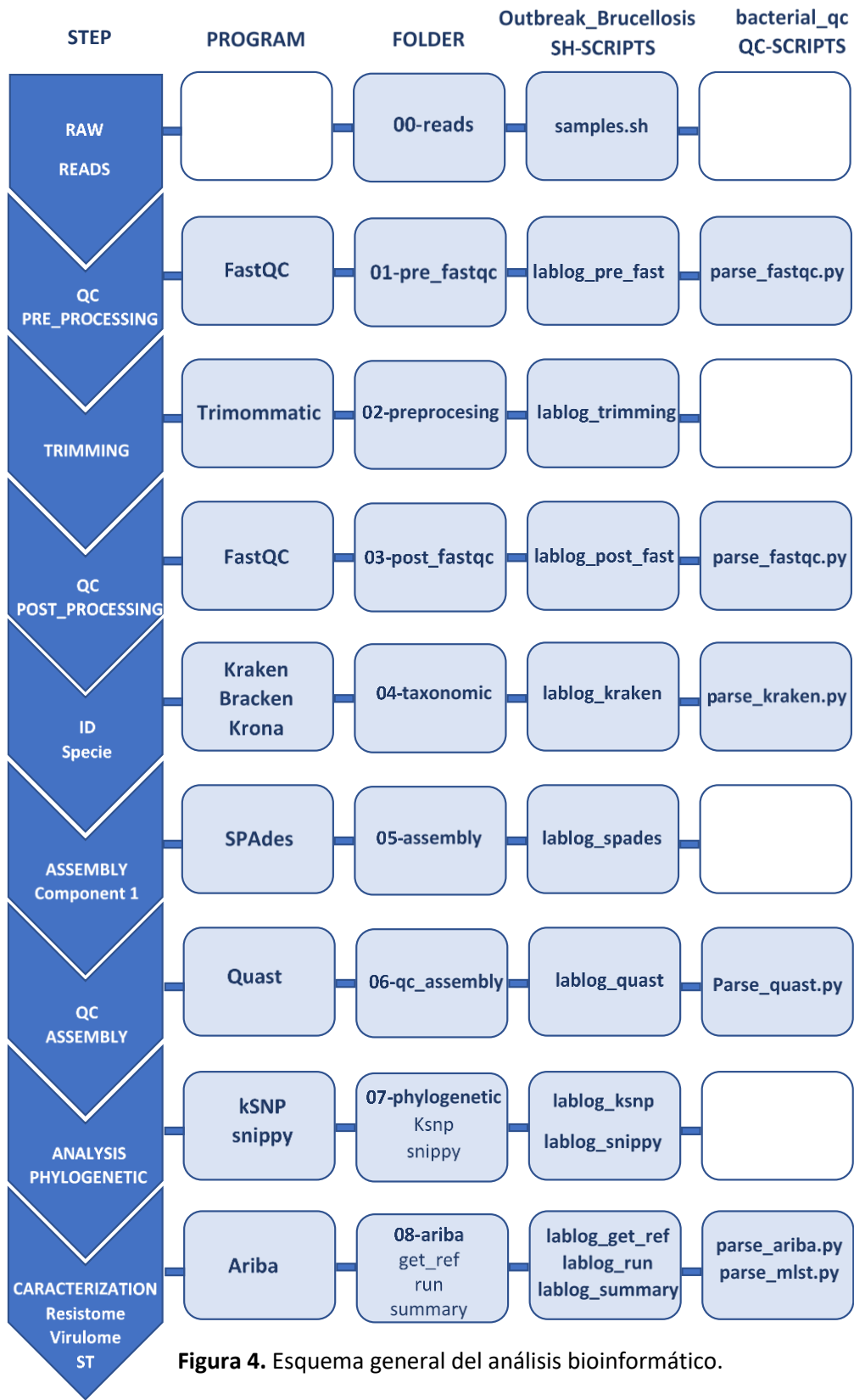
El Componente 1 está formado por las lecturas Illumina MiSeq paired-end ( $2 \times 150$  bp) originarias de un total de 260 genomas de *Brucella spp.* de las especies *B. abortus*, *B. canis*, *B. melitensis* y *B. suis*, 60 de estos genomas completos fueron aislados clínicos, 35 aislados alimentarios, 82 aislados animales y 83 aislados ambientales.

Por otro lado, el Componente 2 está formado por lecturas Illumina MiSeq paired-end ( $2 \times 250$ bp) procedentes de 14 datos metagenómicos de muestras de alimentos.



**4.3 Desarrollo del flujograma de trabajo propuesto para el análisis genómico y metagenómico del estudio.**

En la figura 4 se detallan los diferentes programas, carpetas y scripts utilizados en cada una de las etapas del análisis bioinformático desarrollado en este estudio.



**Figura 4.** Esquema general del análisis bioinformático.

Los scripts utilizados para el análisis de este estudio que permiten la automatización de procesos en la plataforma Linux (SH-SCRIPTS) se encuentran publicados en repositorio público de GitHub [https://github.com/sgonzalezbodi/Outbreak\\_Brucellosis](https://github.com/sgonzalezbodi/Outbreak_Brucellosis). Asimismo, los scripts de Python para la recopilación de los datos más relevantes de cada etapa, control de calidad y análisis descriptivo (QC-SCRIPTS) se encuentran publicados en la plataforma GitHub de la Unidad de Bioinformática del Instituto de Salud Carlos III, [https://github.com/BU-ISCI/bacterial\\_qc](https://github.com/BU-ISCI/bacterial_qc), publicado como paquete de control de calidad desarrollado durante las prácticas extracurriculares del máster.

#### 4.3.1 Datos crudos y listado de muestras

Inicialmente los datos crudos recibidos fueron almacenados en la carpeta 00-reads creando a su vez dos subcarpetas, una para cada Componente. Seguidamente se creó un documento de texto donde se listan las muestras de ambos Componentes que fue utilizado para ejecutar los diferentes comandos de cada etapa de forma conjunta. Sample.sh fue el script desarrollado para la ejecución de esta tarea.

#### 4.3.2 Control y filtrado por calidad

En ambos Componentes se recibieron los archivos FASTQ para cada uno de los paired-ends (uno para el forward o secuencia sentido y otro para el reverse o secuencia antisentido) correspondientes a cada muestra. Dicho archivo contiene las lecturas y los datos de calidad (phred score) basados en código ASCII (en el que cada letra representa un valor numérico), de manera que una secuencia con phred score=10 tiene un 90% de eficacia (una base mal secuenciada cada 10) y una secuencia con phred score=30 tiene un 99,9% de eficacia (una base mal secuenciada cada 1.000).

Inicialmente se analizó la calidad de los ficheros crudos FASTQ mediante el programa FastQC (versión 0.11.9)<sup>16</sup>. Este análisis permite detectar posibles contaminaciones de las librerías, errores sistemáticos en la secuenciación, presencia de barcodes y primers de secuenciación que nos ayuda a tomar decisiones a posterior para el procesamiento y filtrado de las lecturas crudas a partir de la información representada en los distintos módulos de análisis de FastQC.

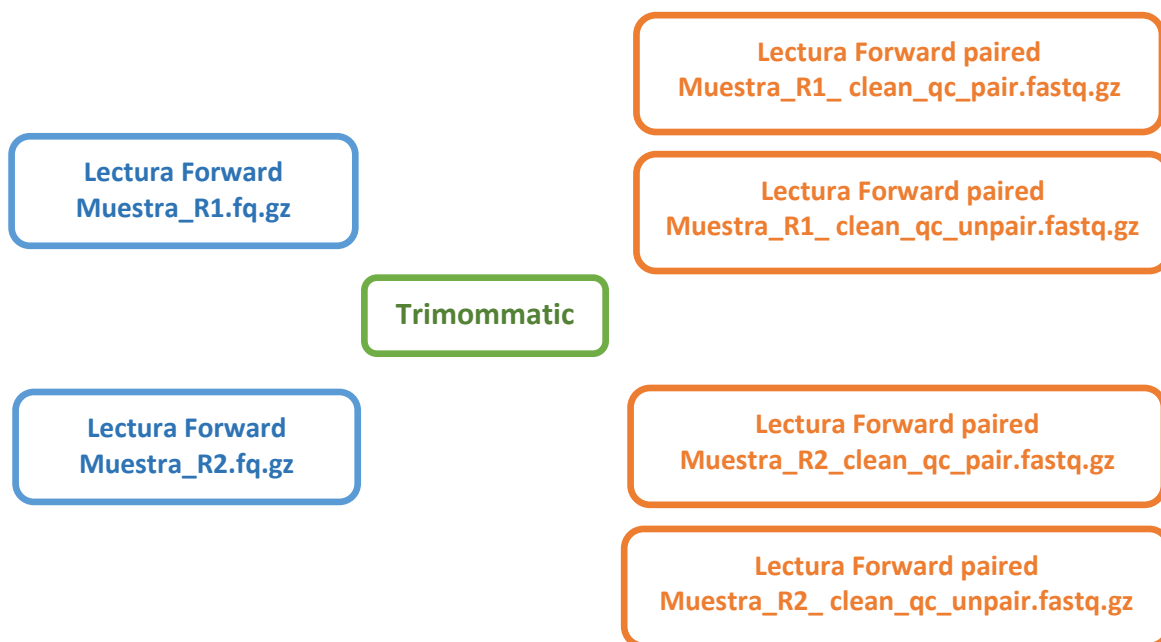
Estos módulos corresponden a: estadísticas básicas de las lecturas; calidad promedio de las bases en cada posición de las lecturas; distribución de la calidad de las lecturas; contenido de bases nucleotídicas por posición; contenido de GC promedio de las lecturas; contenido de GC medio por base; contenido de "N" por base; distribución de longitud de las lecturas; y secuencias duplicadas.

Tras realizar el análisis de las lecturas crudas, de acuerdo con los valores de calidad obtenidos en la etapa previa se determina la realización de un filtrado de estas lecturas para depurarlas y prepararlas para los análisis posteriores utilizando el programa

<sup>16</sup> Andrews S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. 2010. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

Trimmomatic (versión 0.39-1)<sup>17</sup>. Los parámetros de calidad fueron: escaneo de las lecturas con un slidingwindow de 4 pb de ancho, cortando cuando la calidad media de la ventana de la lectura caía debajo del valor de Phred de 15. Las lecturas con longitud menor de 60 fueron descartadas.

Además, al tratarse de lecturas Illumina paired-end el programa verifica que cada par de lecturas esté completo. Las lecturas que no tienen su par o a las que su par fue eliminado por no cumplir alguno de los criterios de calidad son separadas en un archivo diferente calificados como unpaired. Los archivos producidos durante esta etapa de filtrado se representan en la figura 5.



**Figura 5.** Esquema del procesamiento de lecturas en modo paired-end. Trimmomatic.

Finalmente, se repetirá el análisis de calidad de las secuencias filtradas con el programa FastQC con el objetivo de evaluar la calidad del proceso de filtrado realizado. Se utilizará la herramienta MultiQC (versión 1.7) para visualizar todas las muestras de forma conjunta<sup>18</sup>.

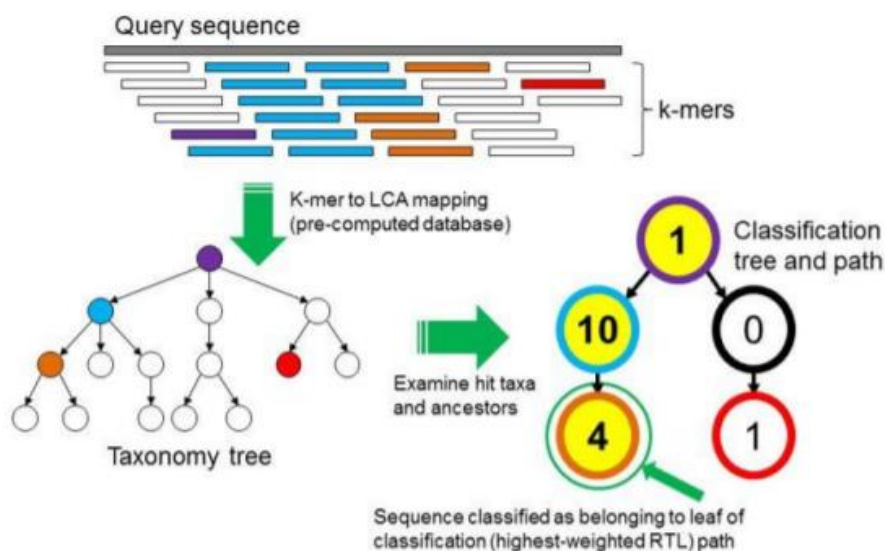
Además, se extraerán los datos obtenidos con el script parse\_fastqc.py para la elaboración de una tabla que nos permite calcular y comparar los diferentes estadísticos de interés como métricas de calidad de la secuenciación, como son la cantidad de lecturas, longitud promedio de las lecturas, rango de longitud de las lecturas, porcentaje de bases nucleotídicas GC. Este primer apartado comprenderá las carpetas: 01-pre\_fastqc, 02-preprocessing, 03-post\_fastqc.

<sup>17</sup> Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

<sup>18</sup> Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047–8.

### 4.3.3 Identificación de especies bacterianas - Análisis taxonómico

Para la identificación de especies bacterianas presentes en las muestras de ambos Componentes se utilizó la herramienta Kraken (versión 2.0.8)<sup>19</sup>. Esta herramienta permite identificar taxonómicamente una especie o género de forma rápida sin necesidad de ensamblar las lecturas, utilizando alineamientos exactos de k-mers contra su propia base de datos resultando muy útil para detectar poblaciones microbianas en datos metagenómicos. Asimismo, supone una importante herramienta de calidad para detectar posibles contaminaciones cuando se secuencian genomas completos únicos. Inicialmente, todos los k-mers de las lecturas de cada muestra se mapean de manera independiente frente al ancestro común más próximo. Seguidamente se realiza una clasificación en forma de árbol en la que los diferentes nodos corresponden con un taxón, siendo elegida como ruta de clasificación de la lectura aquella con mayor número de k-mers mapeados. En la figura 6 se representa el algoritmo de clasificación de Kraken.



**Figura 6.** Representación del algoritmo utilizado en la herramienta Kraken.

La base de datos utilizada es la minikraken2\_v2\_8GB\_201904\_UPDATE que contienen genomas de bacterias, arqueas, virus y hongos. Para la presentación de los resultados del análisis la herramienta utilizará el esquema taxonómico del NCBI. En la figura 6 se presenta un ejemplo del informe de resultados producidos por Kraken.

<sup>19</sup> Wood DE and Steven L Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 2014, 15:R46.

86.90	938804	938804	U	0	unclassified
13.10	141466	44	R	1	root
13.08	141259	112	R1	131567	cellular organisms
8.47	91523	41	D	2	Bacteria
8.32	89853	8565	D1	1783272	Terrabacteria group
4.27	46117	0	P	201174	Actinobacteria
4.27	46099	1	C	1760	Actinobacteria
4.25	45924	0	O	85010	Pseudonocardiales
4.25	45924	0	F	2070	Pseudonocardiaceae
4.25	45904	0	G	65496	Actinoalloteichus
4.25	45904	45904	S	2072503	Actinoalloteichus sp. AHMU CJ021
0.00	6	0	G	1847	Pseudonocardia
0.00	4	4	S	1690815	Pseudonocardia sp. HH130630-07
0.00	2	0	S	240495	Pseudonocardia dioxanivorans
0.00	2	2	S1	675635	Pseudonocardia dioxanivorans CB1190
0.00	4	0	G	1851	Saccharomonospora
0.00	2	0	S	40988	Saccharomonospora azurea
0.00	2	2	S1	882081	Saccharomonospora azurea NA-128
0.00	1	0	S	40989	Saccharomonospora cyanea
0.00	1	1	S1	882082	Saccharomonospora cyanea NA-134
0.00	1	0	S	40990	Saccharomonospora glauca

**Figura 7.** Ejemplo de informe de resultados de Kraken. Se trata de un archivo delimitado por tabulaciones, con una línea por taxón. Los campos de la salida, de izquierda a derecha son: Porcentaje de lecturas del clado en este taxón, número de lecturas totales del clado en este taxón, número de lecturas asignadas específicamente a este taxón, rango de código de letras que indican (U) no clasificado, (D) dominio, (K) Reino, (P) Filo / División, (C) clase, (O) orden, (F) Familia, (G) genero, o (S) especie, NCBI ID taxonómico (Base de datos Taxonomy; <https://www.ncbi.nlm.nih.gov/taxonomy>), descripción.

Seguidamente el output producido por Kraken se reanalizará con la herramienta Bracken (versión 2.5.0)<sup>20</sup> que permite obtener la abundancia de cada especie detectada utilizando la reestimación Bayesiana de abundancia. Un ejemplo del documento producido tras el análisis con Bracken se representa en la tabla 3.

**Tabla 3.** Ejemplo de informe de resultados de Bracken. Los campos de la salida, de izquierda a derecha son: descripción, NCBI ID taxonómico, Rango de código de letras que indican (U) no clasificado, (D) dominio, (K) Reino, (P) Filo / División, (C) clase, (O) orden, (F) Familia, (G) genero, o (S) especie, lecturas asignadas por Kraken2, lecturas añadidas, nuevas lecturas asignadas, dato de abundancia.

name	taxonom y_id	taxonomy _lvl	kraken_assigned _reads	added_re ads	new_est_r eads	fraction_total_ reads
<b>Brucella melitensis</b>	29459	S	55198	293494	348692	0.62276
<b>Homo sapiens</b>	9606	S	63379	62	63441	0.11331
<b>Actinoalloteichus sp. AHMU CJ021</b>	2072503	S	56206	644	56850	0.10153
<b>Bacillus cereus</b>	1396	S	29436	8757	38193	0.06821
<b>Brucella abortus</b>	235	S	317	14525	14842	0.02651
<b>Brucella canis</b>	36855	S	65	12162	12227	0.02184
<b>Lactococcus lactis</b>	1358	S	6260	1180	7440	0.01329
<b>Brucella suis</b>	29461	S	271	5041	5312	0.00949
<b>Lactobacillus helveticus</b>	1587	S	1989	616	2605	0.00465
<b>Brucella pinnipedialis</b>	120576	S	74	1462	1536	0.00274
<b>Brucella ovis</b>	236	S	222	628	850	0.00152

<sup>20</sup> Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in meta-genomics data. PeerJ Computer Science, 2017, 3:e104.

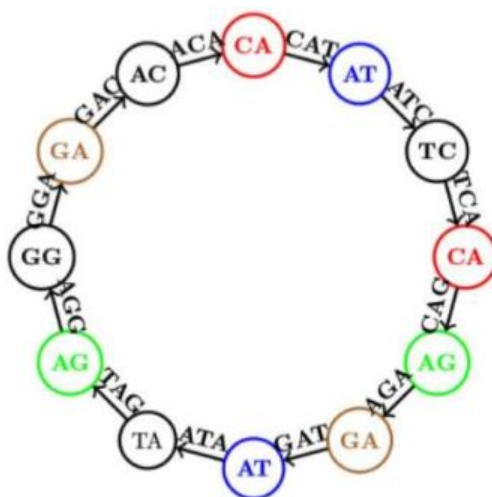
Además, se desarrolló el script `parse_kraken.py` que permite unificar los resultados del conjunto de muestras de cada Componente y de esta forma poder analizar los resultados obtenidos del análisis taxonómico en un único archivo que contiene la información de los 2 hits mayoritarios.

Por último, se utilizó la herramienta Krona (versión 2.7.1)<sup>21</sup> para la visualización interactiva del contenido de taxones presentes en las muestras del Componente 2. Este apartado comprenderá la carpeta: 04-taxonomic\_analysis.

#### 4.3.4 Ensamblado

El ensamblado es el proceso mediante el cual un gran número de secuencias cortas se unen en fragmentos mayores llamados contigs o scaffolds que intentan representar los genomas originales a partir de los cuales se originaron estas. Existen dos modos para realizar esta tarea: alineamientos basados en una referencia y ensamblados de novo donde no se utiliza ninguna referencia y se basan en el solapamiento de las distintas secuencias cortas.

En este trabajo se realizó el ensamblado de novo de las lecturas filtradas por calidad procedentes del directorio 02-preprocessing de los genomas completos del Componente 1 mediante el ensamblador SPAdes (versión 3.13.0)<sup>22</sup>. Este programa está especializado en ensamblar genomas de organismos unicelulares y utiliza para el solapamiento entre lecturas los grafos De-Bruijn (GDB) donde las lecturas son particionadas en subsecuencias de tamaño “k” que se solapan formando una colección de k-mers. En el grafo cada borde es un k-mer que conecta dos nodos, los nodos están formados por el k-mer previo y el k-mer posterior corriéndose una letra hacia atrás (-1) y hacia adelante (+1) del kmer del borde, respectivamente. En la figura 7 se representa el solapamiento de lecturas utilizando los GDB.



**Figura 8.** Representación de Grafo De-Bruijn.

<sup>21</sup> Ondov, B.D., Bergman, N.H. & Phillippy, A.M. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12, 385 (2011). <https://doi.org/10.1186/1471-2105-12-385>.

<sup>22</sup> Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-477.

La longitud de kmer podrá ser introducida por el usuario o el mismo programa puede determinar el valor óptimo. Suele ser adecuado coger una longitud que sea aproximadamente la mitad del tamaño de la secuencia. Los parámetros que se utilizaron para el ensamblaje fueron los predeterminados, con valores de K-mer iguales a 21, 33 y 55.

Una vez realizado el ensamblado se procedió a analizar la eficiencia con el programa QUAST (versión 5.0.2)<sup>23</sup> a partir de los ficheros contigs.fasta procedentes de SPAdes. Como parámetros más importantes a la hora de evaluar la eficacia del ensamblado se encuentran descritos en la tabla 4.

**Tabla 4.** Descripción de los parámetros de más relevantes de QUAST.

Parámetro	Descripción
N50	Corresponde al tamaño del contig situado en el ranking donde ya se acumula el 50% de todas las bases del genoma.
L50	Corresponde al menor número de contigs cuya longitud total representa al menos la mitad de la longitud del ensamblaje.
Número de contigs	El número de contigs que forman el genoma en base a su tamaño, de más grande a más pequeño.

A partir de los resultados obtenidos se extraerán los datos más relevantes con el script `parse_quast.py` para la elaboración de una tabla que nos permita estudiar los diferentes parámetros de calidad previamente descritos. Este apartado comprenderá las carpetas 05-assembly y 06-qc\_assembly.

#### 4.3.5 Análisis filogenético de los aislamientos vinculados epidemiológicamente mediante análisis de polimorfismos de nucleótido único (SNPs).

El análisis filogenético de este estudio fue dirigido a determinar relaciones filogenéticas e identificar aquellas muestras del Componente 1 relacionadas genéticamente y que se vincularon al brote alimentario. Inicialmente para el estudio de las secuencias genómicas completas del Componente 1 se realizó el análisis de polimorfismos únicos con la herramienta kSNP (versión 3.01)<sup>24</sup>. A partir de los resultados de estos análisis se identificaron las muestras del Componente 1 posiblemente vinculadas al brote. Estas muestras junto con aquellas muestras del Componente 2 que tras el análisis taxonómico con Kraken se identificaron con la misma especie vinculada al brote se analizaron con el programa Snippy (versión 4.6.0)<sup>25</sup> para un análisis más profundo de muestras genéticamente muy similares.

<sup>23</sup> Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*. 2018;34(13):i142–50.

<sup>24</sup> Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics*. 2015;31(17):2877–2878.

<sup>25</sup> Seemann T. Snippy: fast bacterial variant calling from NGS reads. 2015. Available from: <https://github.com/tseemann/snippy>.

## kSNP

kSNP identifica SNPs core (SNPs presentes en todos los genomas del estudio) basado en el análisis k-mers sin utilizar genoma de referencia a partir de los ensamblados obtenidos anteriormente con SPAdes (contigs.fasta). Este programa requiere de una validación previa del k-valor óptimo para el tamaño del del k-mer, que es calculado previamente por la herramienta Kchooser a partir de un archivo multifasta creado con los ensamblados. El comando -core utilizado indica el análisis únicamente para el genoma conservado. Además, con el fin de contar los k-mers en los genomas, kSNP utiliza el software jellyfish (versión 2.0)<sup>26</sup>, luego compara estos del k-mers en todos los genomas para encontrar SNPs. Finalmente crea una matriz de SNPs que se utilizó para reconstruir un árbol filogenético de máxima verosimilitud (ML) utilizando RAxML (versión 8.2.12)<sup>27</sup> con un modelo de sustitución de nucleótidos reversible en el tiempo general (GTR), con 100 repeticiones de Bootstrap. Éste el método de referencia que ha mostrado ser el que mejor bondad tiene en la estimación de árboles filogenéticos, se basa en obtener el árbol que presenta la probabilidad o verosimilitud más alta de producir la alineación observada. De este modo, se construye un árbol basado en una matriz de distancias genéticas donde la longitud de cada rama es proporcional al número de cambios producidos entre las secuencias que se comparan.

## Snippy

Para una alta discriminación entre aislados genéticamente muy similares se utilizó el pipeline de Snippy que mapea las lecturas filtradas contra un genoma de referencia con el objeto de detectar polimorfismos con muy alta sensibilidad.

Inicialmente Snippy utiliza BWA-MEM (versión 0.7.12)<sup>28</sup> para mapeo de lecturas cortas. Los SNP se llamaron usando Freebayes (versión 1.0.2)<sup>29</sup>, que requiere una cobertura de lectura mínima de 10x y una concordancia de lectura del 90% en un lugar para que se informe un SNP. Snippy genera una alineación SNP central, así como una alineación SNP genómica completa. Seguidamente, a partir del fichero coreAln, se utilizó el programa Gubbins (versión 2.3.2)<sup>30</sup> para predecir y eliminar los posibles efectos de recombinación. Este programa utiliza un algoritmo que utiliza estadísticas de escaneo espacial para identificar regiones que contienen densidades elevadas de sustituciones de bases que se vinculan con transferencia de secuencias horizontal entre bacterias.

<sup>26</sup> Guillaume Marcais and Carl Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* (2011) 27(6): 764-770.

<sup>27</sup> Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312-1313. doi:10.1093/bioinformatics/btu033.

<sup>28</sup> Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303:3997. [q-bio]. <https://arxiv.org/abs/1303.3997>.

<sup>29</sup> Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv* 1207.3907. [q-bio.GN]. <https://arxiv.org/abs/1207.3907v2>.

<sup>30</sup> Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 2015;43(3):e15.



Por último, el archivo generado por Gubbins se utilizó como input para crear un árbol de máxima verosimilitud con el programa IQ-Tree (versión 2.0.3)<sup>31</sup>. Se seleccionaron los modelos de sustitución de ADN más adecuados para el cálculo por ModelFinder<sup>32</sup>, incluido en IQ-Tree, incorporando 1000 réplicas para el cálculo del bootstrap ultraparamétrico<sup>33</sup> y 1000 réplicas de bootstrap para el SH-like approximate likelihood ratio test<sup>34</sup>.

Este apartado comprenderá la carpeta 05-phylogenetic que contienen las subcarpetas kSNP y Snippy respectivamente.

#### 4.3.5 Caracterización de los aislados y muestras metagenómicas vinculadas al brote de Brucelosis.

Para la caracterización de las cepas del brote se utilizó el programa ARIBA (versión 2.14.5)<sup>35</sup> que permite detectar la presencia tanto de genes de resistencia a antibióticos, factores de virulencia o replicones plasmídicos a partir de las lecturas filtradas. Este programa utiliza la herramienta BLAST del NCBI (<http://blast.ncbi.nlm.nih.gov>) para buscar en el genoma genes de resistencia a antibióticos que se encuentren en bases de datos como ResFinder (acquired antimicrobial resistance gene finder)<sup>36</sup>, Comprehensive Antimicrobial Resistance Database (CARD)<sup>37</sup>, Antibiotic Resistance Genes Database (ARDB) o Antibiotic Resistance Gene-ANNOTation (ARG-ANNOT)<sup>38</sup>. Para la búsqueda de motivos plasmídicos se utilizó la base de datos no redundante de replicones (secuencias que activan o controlan la replicación del plásmido y son identificativas de un determinado tipo de plásmido) Plasmidfinder<sup>39</sup>, que asigna a un determinado grupo Inc (grupo de incompatibilidad plasmídica) o Rep. En cuanto a los genes de virulencia, se utilizó la base de datos denominada Virulence Factor Database (VFDB)<sup>40</sup>, que contiene los principales determinantes de virulencia bacterianos.

<sup>31</sup> Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268-274. doi:10.1093/molbev/msu300.

<sup>32</sup> Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587-589.

<sup>33</sup> Minh, B. Q., Nguyen, M. A. T. y von Haeseler, A. (2013) «Ultrafast Approximation for Phylogenetic Bootstrap», *Molecular Biology and Evolution.* Oxford University Press, 30(5), pp. 1188-1195.

<sup>34</sup> Guindon, S. et al. (2010) «New Algorithms and Methods to Estimate Maximum Likelihood Phylogenies: Assessing the Performance of PhyML 3.0», *Systematic Biology.* Oxford University Press, 59(3), pp. 307-321.

<sup>35</sup> Hunt M, Mather AE, Sánchez-Busó L, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom.* 2017;3(10):e000131. Published 2017 Sep 4. doi:10.1099/mgen.0.000131.

<sup>36</sup> Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother.* 2012;67(11):2640-2644. doi:10.1093/jac/dks261

<sup>37</sup> McArthur AG, Waglechner N, Nizam F, et al. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother.* 2013;57(7):3348-3357.

<sup>38</sup> Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother.* 2014;58(1):212-220.

<sup>39</sup> A. Carattoli, E. Zankari, A. García-Fernández, M. Voldby Larsen, O. Lund, L. Villa, F. Møller Aarestrup, H. Hasman. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother.* 58 (2014), pp. 3895-3903

<sup>40</sup> Chen L, Yang J, Yu J, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 2005;33(Database issue):D325-D328.

El pipeline de Ariba se divide en diferentes etapas. Inicialmente se descargan las bases de datos de interés, seguidamente se realiza un mapeo donde de forma individual se obtienen los resultados de cada muestra. Finalmente, con el comando summary se obtiene un archivo csv donde se engloban los resultados de todas las muestras analizadas. Este apartado comprenderá la carpeta: 08-ariba.

Asimismo, se realizó la tipificación multilocus de secuencias (*Multi-locus Sequence Typing* (MLST)) de todas las cepas del Componente 1 y aquellas vinculadas al brote del Componente 2. Este apartado comprenderá la carpeta: 08-ariba\_mlst.

Esta metodología está basada en la secuenciación de fragmentos de nueve genes conservados “housekeeping” que no se encuentran sometidos a presión selectiva<sup>41</sup>. Siguiendo el esquema descrito por Whatmore y colaboradores <sup>42</sup> publicado en la base datos <https://pubmlst.org/brucella/>, a cada variación alélica de cada uno de los genes se le asignó un número, y a cada combinación de 9 alelos (perfil alélico) se le asignó una secuencia tipo (ST) (Tabla 5).

**Tabla 5.** Genes de mantenimiento utilizados para la tipificación por MLST.

Locus	Putative function
gap	glyceraldehyde 3-phosphate dehydrogenase
aroA	3-phosphoshikimate1-carboxyvinyltransferase
glk	glucokinase
dnaK	chaperone protein
gyrB	DNA gyrase B subunit
trpE	anthranilate synthase
omp25	25 kDa outer-membrane protein
cobQ	cobyric acid synthase
int-hyp	upstream and extreme 5' of hypothetical protein

#### 4.3.6 Visualización de los resultados

Los ficheros en formato NEWICK obtenido de RAxML y IQ-Tree se cargaron en la plataforma Microreact4 (<https://microreact.org/showcase>) para la visualización integrada de los conjuntos de datos de secuenciación y otras variables de interés como ST, Componente, tipo de muestra etc.

<sup>41</sup> Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences.* 1998;95(6):3140-145.

<sup>42</sup> Whatmore AM, Perrett LL, MacMillan AP. Characterisation of the genetic diversity of *Brucella* by multilocus sequencing. *BMC Microbiol.* 2007;7:34.

## 5. Resultados

### 5.1 Análisis de calidad de las lecturas filtradas (Ficheros FastQC)

A continuación, se representan las figuras y datos más relevantes de los análisis de calidad de ambos componentes utilizando el programa MultiQC y los datos descriptivos obtenidos a partir de la tabla de resultados del script `parse_fastqc.py`. En la Tabla 1 de los anexos se detallan los parámetros de calidad principales de cada una de las muestras.

#### Componente1:

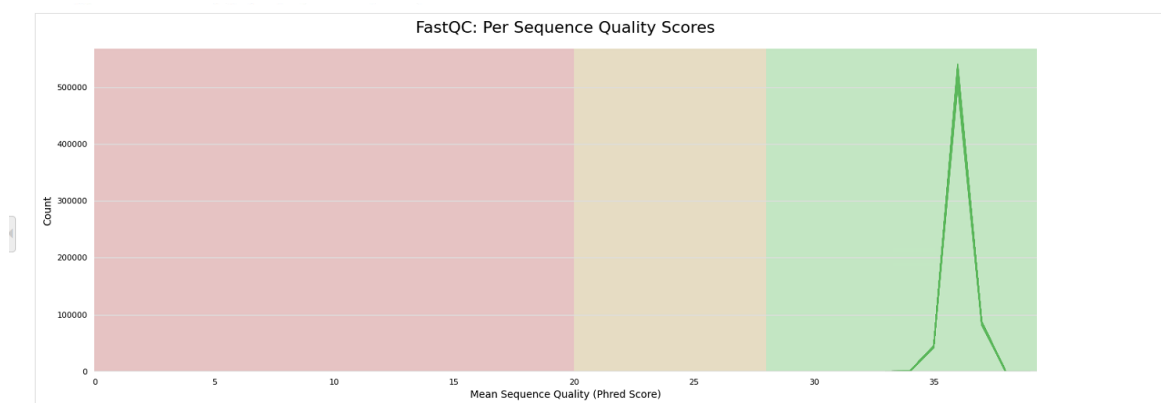
En las Figuras 8 y 9 se representa la calidad de las lecturas filtradas del Componente 1 de acuerdo con la posición de bases generadas (150 bases) y al número de lecturas obtenido ( $658333,5 \pm 5552,1$ ; mínimo 626861; máximo 673704) para las 260 muestras. La división en colores de las figuras representa en todos los casos la medida de la calidad de acuerdo con el parámetro Phred score, distinguiéndose tres zonas:

- **Zona verde:** 28-34. Corresponden a una muy buena calidad.
- **Zona naranja:** zona de calidad intermedia (20-28).
- **Zona roja:** zona de mala calidad (0-20).

En ambas figuras se observa que todas las muestras del Componente 1 tuvieron un valor adecuado de Phred score agrupándose en ambos casos en la zona verde que representa una alta calidad.

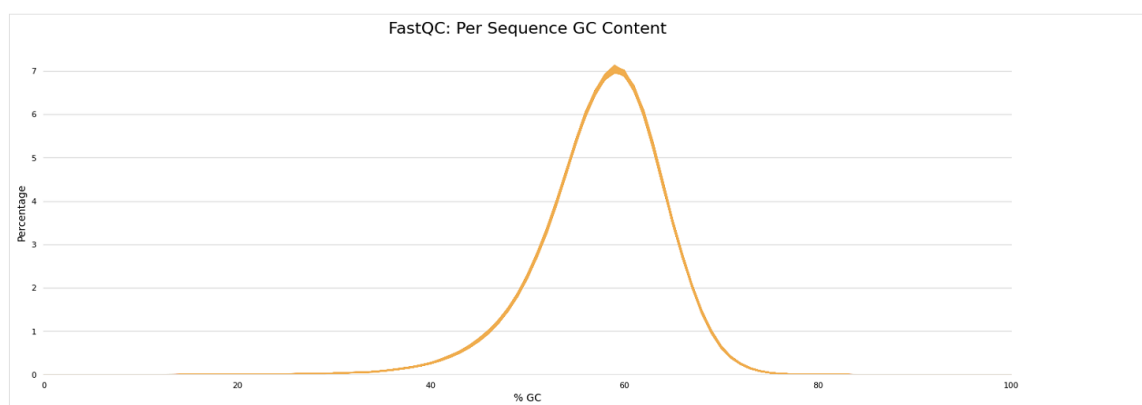


**Figura 9.** Phred score por base



**Figura 10.** Phred score por lecturas.

En relación con el porcentaje medio de nucleótido G y C, la media del %GC fue de  $57 \pm 0\%$  para las 260 muestras (Figura 10).



**Figura 11.** Contenido de GC% por secuencia

### Componente 2:

En la tabla 6 se representan el total de las secuencias en millones, el promedio de longitud de secuencia y contenido medio de GC de cada una de muestras metagenómicas del Componente 2.

Tabla 6. Datos de calidad del Componente 2 visualizados con MultiQC.

Sample Name	% GC	Length	M Seqs
Food_sample01_R1	44%	212 bp	1.1
Food_sample01_R2	44%	193 bp	1.1
Food_sample02_R1	45%	194 bp	2.0
Food_sample02_R2	45%	183 bp	2.0
Food_sample03_R1	44%	220 bp	0.7
Food_sample03_R2	44%	200 bp	0.7
Food_sample04_R1	46%	202 bp	1.7
Food_sample04_R2	46%	188 bp	1.7
Food_sample05_R1	46%	202 bp	1.6
Food_sample05_R2	46%	187 bp	1.6
Food_sample06_R1	46%	203 bp	1.6
Food_sample06_R2	46%	190 bp	1.6
Food_sample07_R1	46%	203 bp	1.7
Food_sample07_R2	46%	190 bp	1.7
Food_sample08_R1	46%	208 bp	2.3
Food_sample08_R2	46%	193 bp	2.3
Food_sample09_R1	46%	204 bp	1.8
Food_sample09_R2	46%	189 bp	1.8
Food_sample10_R1	46%	204 bp	2.0
Food_sample10_R2	46%	189 bp	2.0
Food_sample11_R1	44%	216 bp	1.7
Food_sample11_R2	44%	201 bp	1.7
Food_sample12_R1	46%	204 bp	2.4
Food_sample12_R2	45%	191 bp	2.4
Food_sample13_R1	44%	203 bp	2.7
Food_sample13_R2	44%	185 bp	2.7
Food_sample14_R1	44%	210 bp	1.6
Food_sample14_R2	43%	186 bp	1.6

En las Figuras 11 y 12 se representa la calidad de las lecturas filtradas del Componente 2 de acuerdo con la posición de bases generadas (35-251bases) y al número de lecturas obtenido (1647615,2±518849,7; mínimo 760316; máximo 2831532) para las 14 muestras.

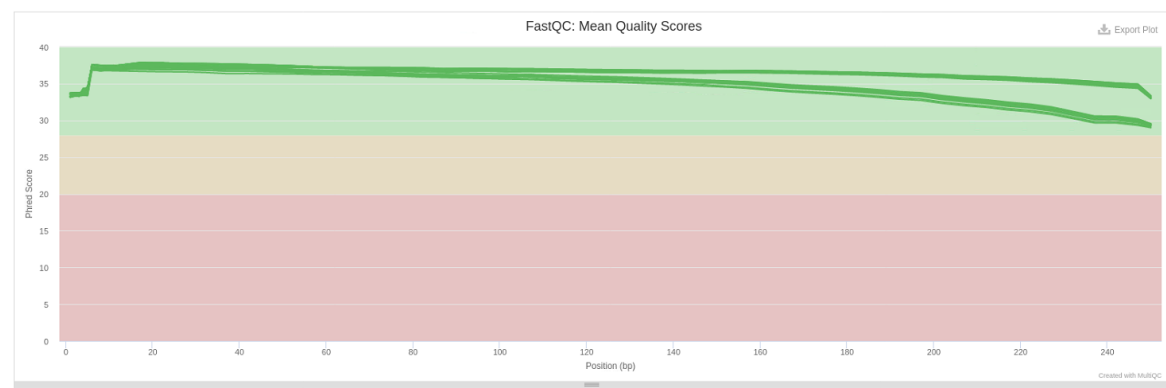
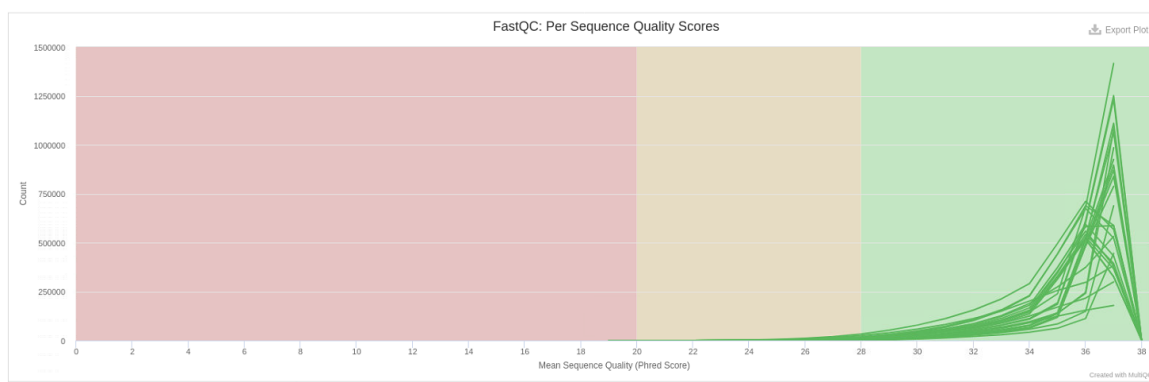
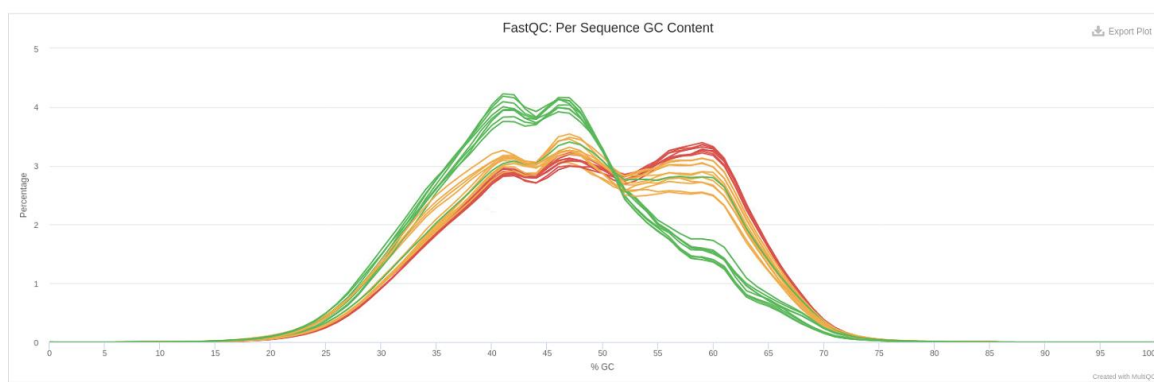


Figura 12. Phred score por base



**Figura 13.** Phred score por lecturas.

En relación al porcentaje medio de nucleótido G y C, el promedio del %GC fue de  $45 \pm 2\% \pm 1,9$ ; mínimo 44; máximo 46 para las 14 muestras (Figura 13).



**Figura 14.** Contenido de GC% por secuencia.

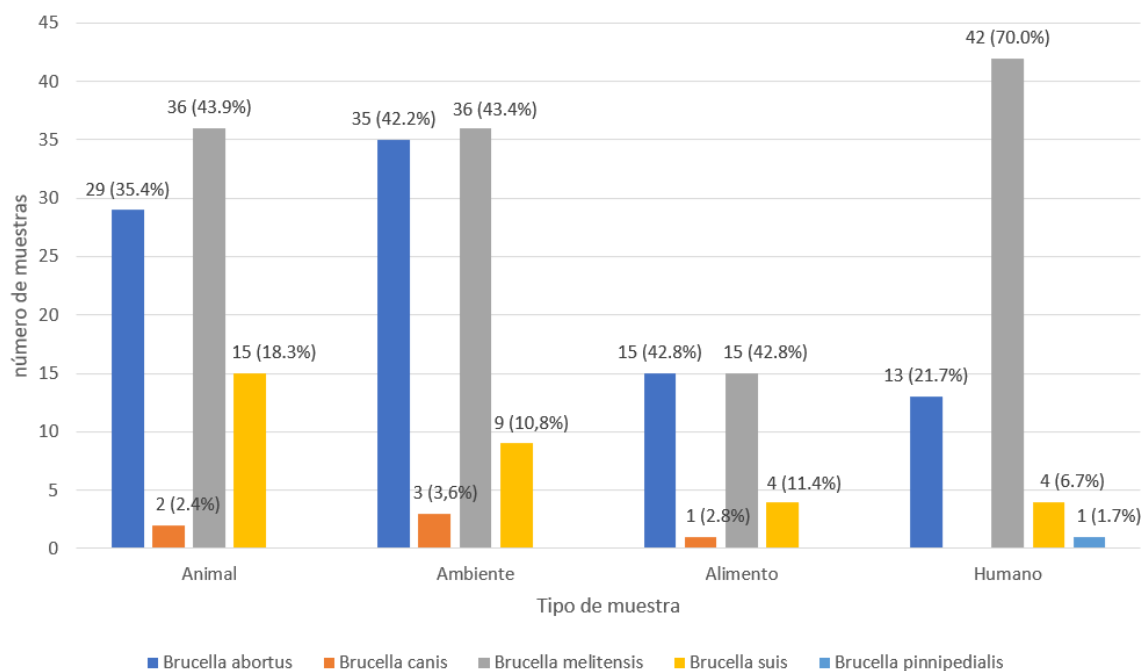
## 5. 2 Análisis taxonómico

En la Tabla 2 de los anexos se detallan los resultados del análisis de la identificación a nivel de especie con Kraken y Bracken para cada una de las muestras del Componente 1. En la tabla aparecen resumidos los parámetros del hit con mayor número de lecturas asignadas utilizando el script `parse_kraken.py`.

El total de las 260 muestras del Componente 1 se distribuyeron en las siguientes especies: 129 (49.6%) *Brucella melitensis*, 92 (36.4%) *Brucella abortus*, 32 (12.3%) *Brucella suis*, 6 (2.3%) *Brucella canis*, 1 (0.4%) *Brucella pinnipedialis*.

La distribución de especies de acuerdo con el tipo de muestra se representa en la figura 14. Para todos los tipos de muestras (animal, ambiente, alimento y humano) la especie predominante fue *Brucella melitensis*.

## Resultados



**Figura 15.** Distribución de especies de acuerdo al tipo de muestra.

En relación al análisis taxonómico de las muestras de metagenómica del Componente 2 los resultados detallados de cada una de las muestras se encuentran descritos en la tabla 3 de anexos obtenidos a partir del script `parse_kraken.py`.

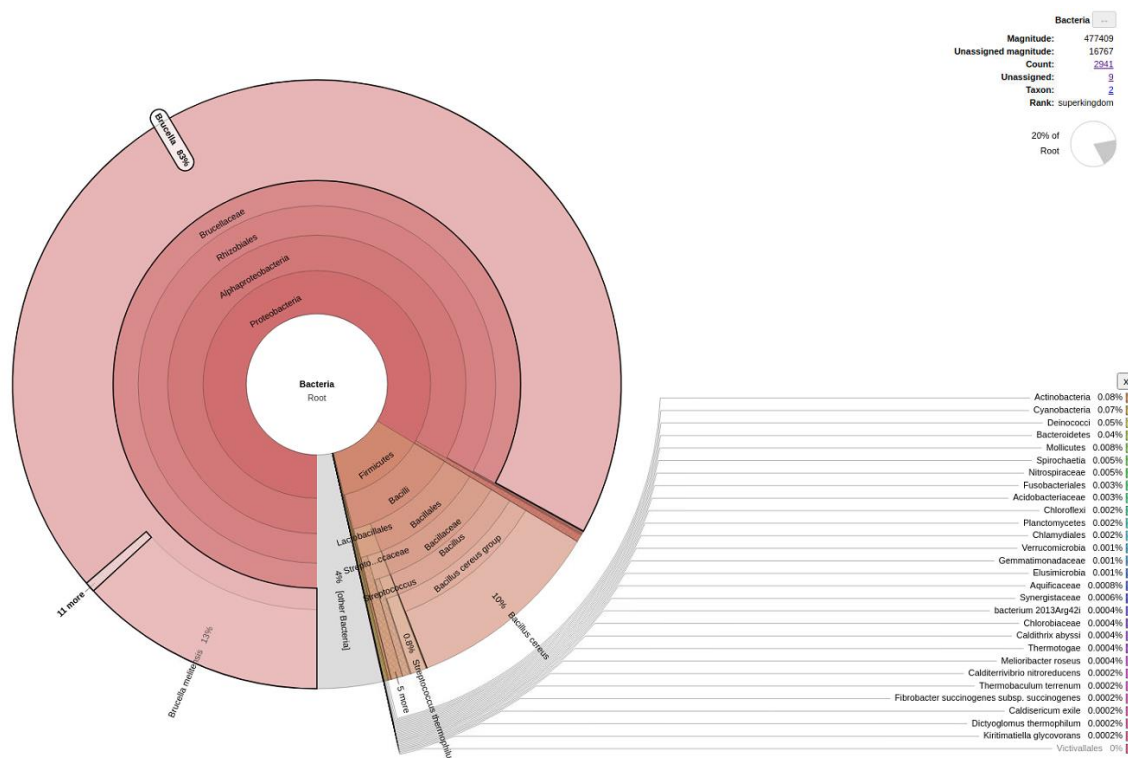
Del total de las 14 muestras, en 10 de ellas se identificaron alguna de las especies de *Brucella spp.* presentes también en el Componente 1 (Tabla 7). Tal y como se observó en el Componente 1 la especie predominante fue *B. melitensis* presente en 4 de las 10 muestras.

**Tabla 7.** Identificación de la especie mayoritaria de las muestras del Componente 2

Nombre de la muestra	Especie asignada con mayor número de lecturas
Food_sample01	<i>Homo sapiens</i>
Food_sample02	<i>B. abortus</i>
Food_sample03	<i>Actinoalloteichus sp. AHMU C1021</i>
Food_sample04	<i>B. melitensis</i>
Food_sample05	<i>B. suis</i>
Food_sample06	<i>B. abortus</i>
Food_sample07	<i>B. canis</i>
Food_sample08	<i>B. suis</i>
Food_sample09	<i>B. abortus</i>
Food_sample10	<i>B. melitensis</i>
Food_sample11	<i>Homo sapiens</i>
Food_sample12	<i>B. melitensis</i>
Food_sample13	<i>B. melitensis</i>
Food_sample14	<i>Homo sapiens</i>

En tabla 4 de anexos se detallan los resultados del análisis taxonómico de Kraken y Bracken para muestras del Componente 2. En la tabla aparecen resumidos los parámetros del best y second hit con mayor número de lecturas asignadas utilizando el script `parse_kraken.py`.

Asimismo, se visualizaron y comprobaron los resultados obtenidos por Kraken con la herramienta interactiva Krona, un ejemplo de este análisis se encuentra en la Figura 15.



**Figura 16.** Ejemplo de resultados de Kraken visualizados en la herramienta interactiva Krona.

### 5.3 Análisis de calidad los ensamblados del Componente 1 (Ficheros FASTA)

En la Tabla 3 de los anexos se detallan los parámetros de calidad principales de los ficheros fasta obtenidos a partir del programa QUAST y del script `parse_quast.py` de cada una de las muestras. Los datos medios de los parámetros de calidad de los ensamblados de todas las muestras se detallan en la tabla 8.

**Tabla 8.** Parámetros de calidad de los ensamblados del Componente 1

Parámetros	Media	Desviación estándar
<b>Nº contigs totales</b>	34.5	28.3
<b>Contig de mayor tamaño</b>	485148.2	106608.0
<b>N50</b>	241536.0	65428.3
<b>N75</b>	146578.3	40026.2
<b>L50</b>	5.9	5.1
<b>L75</b>	10.9	10.8



Se observa que la desviación estándar en todos los parámetros es bastante amplia debido a que un grupo de muestras del Componente 1 presentan un ensamblado de peor calidad (Tabla 9). En la tabla 9 podemos comparar los diferentes parámetros de calidad de una muestra con un ensamblado óptimo sombreada en verde y el grupo de muestras con peor calidad. Un ensamblado óptimo se caracteriza por el menor número de contigs que presentan el mayor tamaño posible para cubrir la mayor parte del genoma. Es decir, tener contigs que sean lo más largos posibles o que el ensamblaje tenga el mayor N50 (que el tamaño para llegar al 50% de la secuencia sea más grande y significará en cierta medida que los contigs son más largos. Asimismo, Quast evaluó que el número total de genes en el conjunto ensamblado sobre el genoma de referencia descartando así posibles missassemblies.

**Tabla 9.** Principales parámetros de calidad de un subgrupo de muestras del Componente 1.

Muestra	Nº contigs totales	Contig de mayor tamaño	N50 <sup>a</sup>	N75 <sup>b</sup>	L50 <sup>c</sup>	L75 <sup>d</sup>
<b>Animal_331</b>	27	419817	222072	159943	6	10
Environment_138	90	243999	85412	46204	12	26
Environment_175	130	176319	48778	28434	22	43
Environment_320	429	48838	11712	6729	79	167
Environment_574	94	144578	64905	38800	19	35
Environment_591	79	320063	144290	74875	7	15
Environment_966	84	173401	78199	38783	15	30
Food_215	116	218428	56816	31409	17	37
Human_496	84	292954	138433	112419	9	15
Human_983	130	156797	48232	32156	24	43

<sup>a</sup>N50 es la longitud para la cual la colección de todos los contigs de esa longitud o más larga cubre al menos la mitad de un ensamblado. <sup>b</sup>N75 se define de manera similar a N50 pero con 75% en lugar de 50%. <sup>c</sup>L50 es el número de contigs igual o mayor que N50 que cubren la mitad del ensamblado. <sup>d</sup>L75 es el número de contigs igual o mayor que N75 que cubren  $\frac{3}{4}$  del ensamblado.

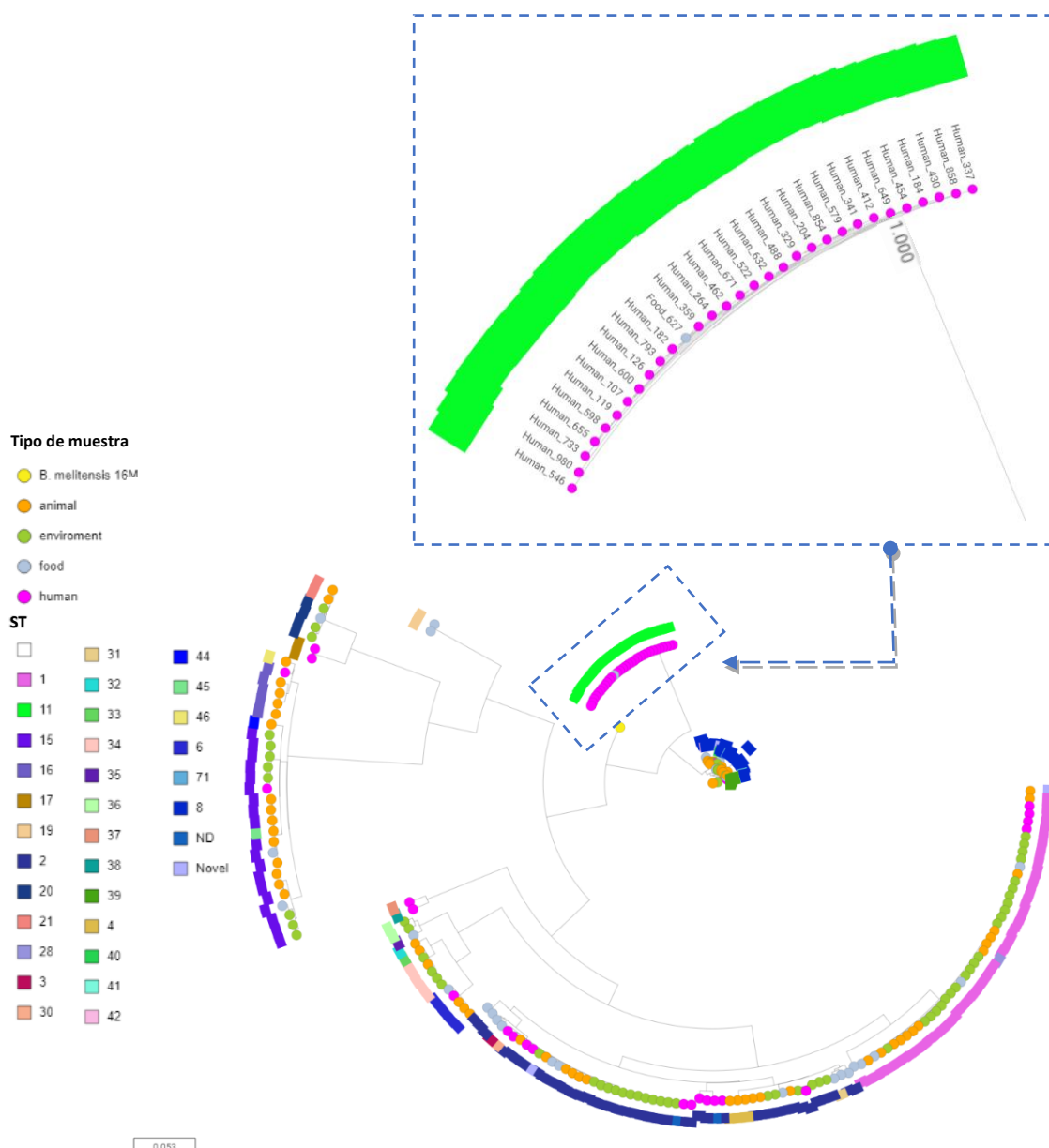
Los datos medios de los parámetros de calidad de los ensamblados excluyendo aquellas muestras que presentan una peor calidad se detallan en la tabla 10. Observamos que la desviación estándar en todos los parámetros se reduce considerablemente al excluir estas muestras del análisis.

**Tabla 10.** Parámetros de calidad de los ensamblados del Componente 1 excluyendo muestras que presentaron un peor ensamblado.

Parámetros	Media	Desviación Estándar
<b>Nº contigs totales</b>	30.8	6.1
<b>Largest contig</b>	495470.8	92014.0
<b>N50</b>	247500.3	57827.8
<b>N75</b>	150201.4	35335.9
<b>L50</b>	5.2	1.1
<b>L75</b>	9.7	2.2

### 5.3 Análisis filogenético

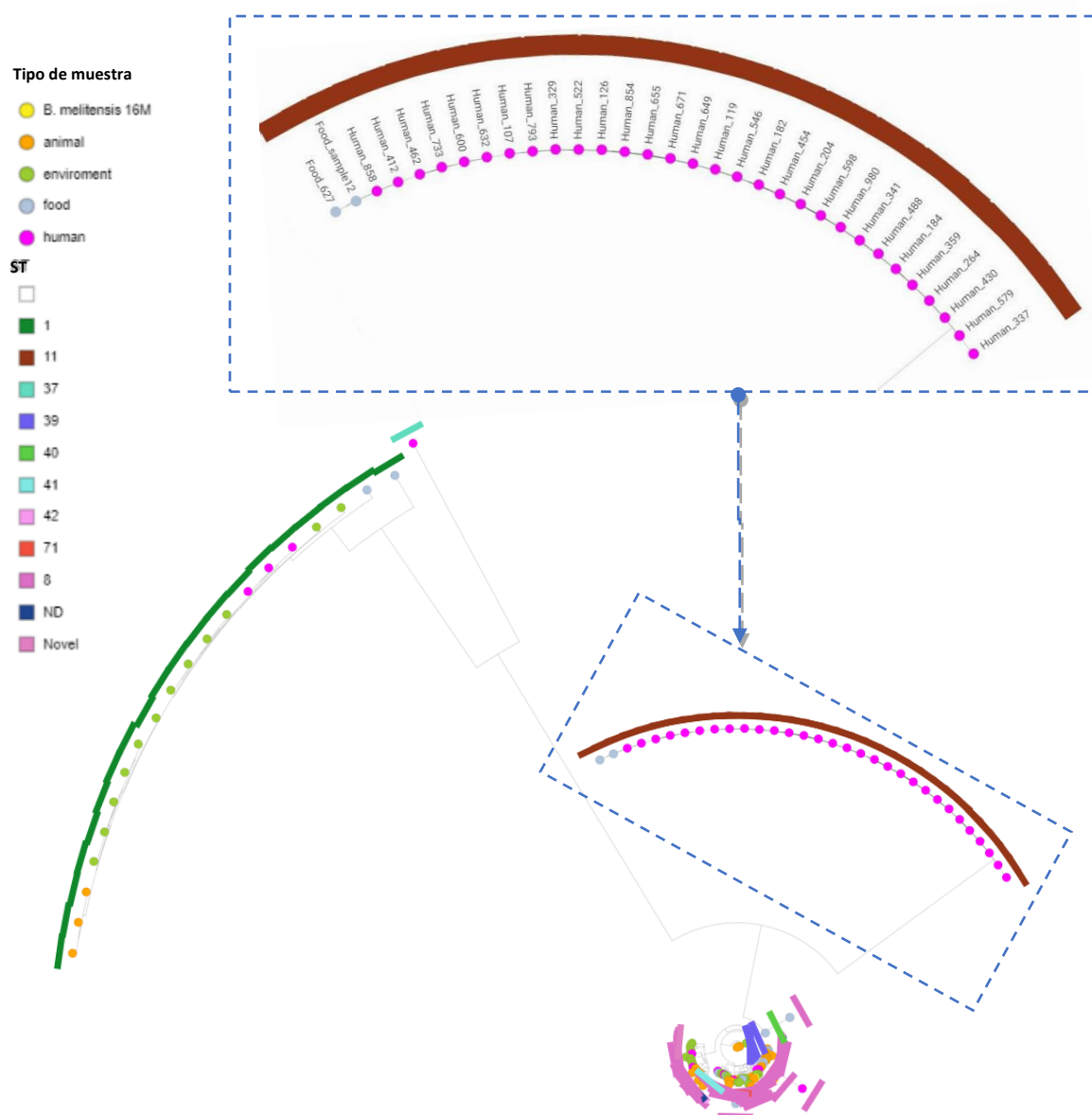
El árbol filogenético kSNP basado en 15383 core SNPs de las 160 genomas del Componente 1 clasificadas según el tipo de muestra y la secuencia tipo (ST) están representados en la Figura 16. Las cepas se agruparon según similitud de STs. De las 43 cepas clínicas de *B. melitensis*, 30 de estas se agruparon en un clado único pertenecientes todas a la ST 11 junto con la muestra de alimento “Food\_627” con un valor de soporte total. Queda patente la baja variabilidad entre las 31 secuencias que conforman este clado que son muy similares entre sí, ya que difieren en aproximadamente en <0,0002% del genoma core con una escala fijada en 0.053 sustituciones por sitio, lo que sugiere un ancestro común que podría deberse a la presencia de un brote.



**Figura 17.** Árbol filogenético kSNP de máxima verosimilitud (ML) generado mediante RaXML a partir de la alineación de los SNPs core generados por kSNP. Se amplía el clado perteneciente a la secuencia tipo 11, donde se agrupan la mayor parte de las muestras de aislados humanos de *B. melitensis*. <https://microreact.org/project/44jDxz3FfWWqGvLJ4TikU5/4c765acc>. Las unidades de la escala son cambios nucleotídicos (SNPs) por posición. Se incluye la información del tipo de

muestra en el extremo de cada rama en forma de puntos y en las circunferencias exteriores la información del ST.

Todas las muestras del Componente 1 identificadas como *B. melitensis* (31) junto con las cuatro muestras metagenómicas del Componente 2 que también fueron identificadas con la misma especie fueron analizadas con el pipeline de Snippy. La cepa de referencia que se utilizó para realizar la inferencia filogenética fue *B. melitensis* biotipo 1 16 M Chr I y Chr II (GeneBank: NC003317, NC003318), ya que es la cepa de referencia representativa de esta especie. El árbol filogenético de Snippy reconstruido con IQ-Tree utilizando el modelo K3Pu+F+R10 a partir de un alineamiento de 6318 SNPs conservados generado en base al mapeo de lecturas filtradas frente a la referencia se representa en la Figura 17.



**Figura 18.** Árbol filogenético SNIPPY reconstruido a partir de un alineamiento de los SNPs conservado generado en base al mapeo de lecturas filtradas frente a la referencia y corregido por recombinación usando Gubbins. <https://microreact.org/project/nN3XnALJMSvn8pBNvhEqwh>. Las unidades de la escala son cambios nucleotídicos (SNPs) por posición. Se incluye la información del

tipo de muestra en el extremo de cada rama en forma de puntos y en las circunferencias exteriores la información del ST.

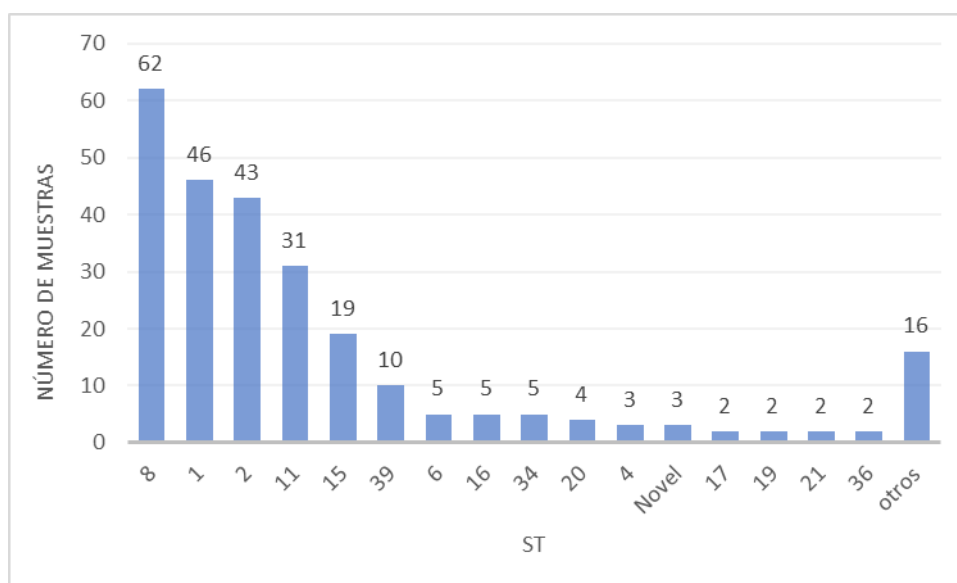
Las cepas se agruparon nuevamente de acuerdo a su STs. Al introducir en el análisis las 4 muestras del Componente 2 identificadas como *B. melitensis*, la muestra “Food\_sample\_12” con ST 11 se agrupó dentro del grupo monofilético sospechoso del brote alimentario formado por las 30 muestras clínicas de *B. melitensis* junto con la muestra de alimento “Food\_627” de la misma ST. Las 32 muestras que conformaron este clado no tenían variaciones y su identidad era del 100%, este dato fue asignado con un valor de soporte total.

#### 5.4 Caracterización de los aislados y muestras metagenómicas vinculadas al brote de Brucelosis.

##### 5.4.1 Tipificación multilocus de secuencias, MLST (ST).

Del análisis de tipificación de las 260 muestras del Componente 1 se asignaron un total de 31 ST ya descritas y 3 ST Novel. El ST predominante fue el ST\_8 que se observó en el 23.85% de las muestras del Componente 1 (Tabla x).

**Tabla 11.** Distribución de STs en las muestras del Componente 1.



Según el tipo de muestra y la especie identificada, se observan diferentes STs predominantes. En la tabla 12 se representa las STs identificadas en los 129 aislados de *B. melitensis* de acuerdo con el tipo de muestra. Tal y como se observa en la tabla 12 la ST11 fue la predominante en las muestras clínicas (30/42;71.4%), la ST8 fue la predominante para los aislados de animal (25/36: 69.4%) y alimento (8/15;53.3%) y la ST1 fue la predominante en los aislados de ambiente (12/36; 33.3%).

**Tabla 12.** Distribución y frecuencias de ST en los aislados de *B.melitensis* (n= 129) de acuerdo al tipo de muestra.

Clínicas (n = 42)			Animal (n=36)			Ambiente (n= 36)			Alimento (n= 15)		
ST	n	%	ST	n	%	ST	n	%	ST	n	%
1	3	7.1	8	25	69.4	1	12	33.3	8	8	53.3
11	30	71.4	1	3	8.3	39	1	2.8	39	3	20.0
37	1	2.4	39	6	16.7	8	22	61.1	1	2	13.3
41	1	2.4	42	1	2.8	Novel	1	2.8	11	1	6.7
8	7	16.7	71	1	2.8				40	1	6.7

Los datos detallados de cada una de las muestras con los alelos de los 9 genes endógenos se encuentran en la tabla 5 de anexos.

#### 5.4.2 Genes de virulencia y resistencia a antibióticos

Los genes de virulencia y resistencia a antibióticos de las 32 cepas causantes del brote alimentario se encuentran en las Tabla 12 y 13 agrupados según su función biológica.

**Tabla 13.** Genes de virulencia en las cepas causantes del brote de Brucelosis.

Función biológica	Gen
Proteínas que suprimen de la respuesta innata	btpA
	btpB
Glucanos cíclicos periplásmicos que permiten la respiración anaeróbica	cgs
Constituyentes del polisacárido de la membrana externa	fabZ
	acpXL
	gmd_2
	htrB_1
	kdsA_1
	kdsB
	lpsA_1
	lpsB_lpcC
	lpxA
	lpxB
	lpxC_1
	lpxD
	lpxE
	manAoAg
	manCoAg
	per
	pgm

	pmm
	ricA
	wbdA
	wbkA
	wbkB
	wbkC
	wboA
	wbpL
	wbpZ
	wzm
	wzt
Sistema de secreción tipo IV virB involucrado en la secreción de moléculas efectoras dentro de la célula hospedadora	virB1
	virB10_2
	virB11
	virB12
	virB2_4
	virB3
	virB4
	virB5
	virB6
	virB7_1
	virB8
	virB9_3

**Tabla 14.** Genes de resistencia a antibióticos en las cepas causantes del brote de Brucelosis.

Función biológica	Gen
Asociado con resistencia a daptomicina	mprF

La llamada a la base de datos PlasmidFinder no encontró ningún resultado.

## 6. Discusión y Perspectivas Futuras

---

Los resultados del análisis taxonómico para la identificación de especies bacterianas y control de calidad con la herramienta Kraken confirmaron que todos los genomas del Componente 1 pertenecían a diferentes especies de *Brucella spp.* y por tanto no existía otras especies contaminantes o minoritarias en los genomas secuenciados. Asimismo, el análisis taxonómico de las diferentes muestras metagenómicas del Componente 2 identificó la población existente en cada una de ellas y la presencia de *Brucella spp.* en 10 de las 14 muestras. La especie identificada mayoritariamente fue *B. melitensis* determinada en 4 de estas muestras.

Los resultados del estudio filogenético realizado con la herramienta kSNP de acuerdo al análisis del genoma basado en SNPs conservados determinaron que la mayor parte de las muestras clínicas se agrupaban en el árbol filogenético en una rama independiente formando un único clado junto con un aislado alimentario “Food\_627”. Todas estas muestras pertenecían a la especie *B. melitensis*. En este procedimiento el alineamiento se produce entre las muestras y no contra un genoma de referencia pudiendo comparar muestras que tengan una mayor diversidad en su genoma. Por tanto, este método, pese a tener un menor poder discriminatorio, se utilizó como screening inicial para clusterizar a las diferentes muestras del estudio y observar su distribución.

En este análisis se detectaron un total de 31 muestras vinculadas al brote de Brucelosis procedentes de aislados de personas enfermas como de alimentos. Seguidamente, conociendo la especie vinculada al brote se realizó un segundo análisis altamente discriminatorio de comparación de SNPs del genoma core con Snippy. Esta herramienta presenta una alta sensibilidad y especificidad ya que utiliza un genoma de referencia para realizar el alineamiento de genomas muy similares. Por esta razón, en este segundo estudio se analizaron sólo los aislados de *B. melitensis* del Componente 1 junto con las muestras metagenómicas que contenían el genoma de esta misma especie. El árbol filogenético obtenido de este análisis ratificó los resultados previamente obtenidos con la herramienta kSNP agrupando nuevamente a los aislados clínicos junto con el aislado alimentario en una rama diferenciada conformando un grupo monofilético determinado por una única cepa. Igualmente, en este mismo clado se agrupó la muestra metagenómica “Food\_sample\_12”. Esto permitió establecer una relación de la muestra metagenómica con la misma cepa del brote y determinar el origen del brote identificando la fuente alimentaria causante del mismo.

Los resultados de la caracterización de las muestras vinculadas al brote con la herramienta Ariba permitió establecer el viruloma y resistoma de la cepa causante del brote. Asimismo, a través de esta misma herramienta se realizó el tipado por MLST determinado que todas las muestras vinculadas al brote pertenecían a la ST11. Asimismo, se tipo al resto de genomas de *Brucella spp.* del Componente 1. Aunque el MLST presentó un buen poder discriminativo, el estudio de los SNPs de cientos de genes conservados permitieron la obtención de conclusiones más robustas.

A la vista de los resultados obtenidos y tomando como referencia los objetivos planteados en este estudio, podemos concluir que las metas establecidas en este proyecto se han desarrollado correctamente. Sin embargo, como futuras líneas de trabajo a introducir en este pipeline de análisis se plantea la implementación del tipificado basado en core genoma MLST (cgMLST) y whole genoma MLST (wgMLST), es decir, basado en el análisis del conjunto genes conservados en todos los genomas utilizando esquemas específicos de especie. De esta forma, poder comparar los resultados obtenidos utilizando esta nueva técnica con el tipado por SNPs ya desarrollado en este estudio.

Las conclusiones obtenidas según los objetivos planteados son:

1. La especie identificada mayoritariamente tanto en el Componente 1 (secuenciación de genomas completos) como en el Componente 2 (secuenciación metagenómica) fue *B. melitensis*
2. Los resultados del estudio filogenético realizado basado en SNPs conservados determinaron que la mayor parte de las muestras clínicas se agrupaban en el árbol filogenético formando un único clado junto con un aislado alimentario "Food\_627".
3. En este análisis se detectaron un total de 31 muestras vinculadas al brote de Brucelosis procedentes de aislados de personas enfermas como de alimentos
4. Se identificó una relación entre la muestra "Food 627" y la muestra metagenómica "Food\_sample\_12", determinando así el origen del brote e identificando la fuente alimentaria causante del mismo.
5. Todas las muestras vinculadas al brote pertenecían a la ST11.
6. La utilización de la epidemiología genómica para resolver este brote de Brucelosis ha permitido discriminar aquellos aislados pertenecientes al mismo y confirmar su dispersión a otras muestras alimentarias
7. Se ha desarrollado un pipeline de análisis de datos, disponible en repositorio público GitHub, que podrá ser utilizado para labores de vigilancia de salud pública en estudios de identificación, seguimiento, evolución y trazabilidad de brotes infecciosos causados por patógenos.



## 8. Anexos

**Tabla Anexo 1.** Parámetros de calidad principales de los ficheros FASTQ del Componente 1 y 2 (parse\_fastqc.py).

Muestra	%GC	Longitud de la secuencia	Secuencias con mala calidad	Secuencias totales
Animal_113	57	150	0	657120
Animal_127	57	150	0	655980
Animal_140	57	150	0	656060
Animal_143	57	150	0	658755
Animal_154	57	150	0	656730
Animal_173	57	150	0	657240
Animal_189	57	150	0	662131
Animal_209	57	150	0	655830
Animal_217	57	150	0	658727
Animal_222	57	150	0	652250
Animal_234	57	150	0	658755
Animal_236	57	150	0	663843
Animal_241	57	150	0	662155
Animal_251	57	150	0	668910
Animal_254	57	150	0	659799
Animal_267	57	150	0	657600
Animal_268	57	150	0	657120
Animal_275	57	150	0	666661
Animal_279	57	150	0	651060
Animal_285	57	150	0	664011
Animal_287	57	150	0	662286
Animal_289	57	150	0	657090
Animal_290	57	150	0	657240
Animal_317	57	150	0	657270
Animal_325	57	150	0	657824
Animal_331	57	150	0	659640
Animal_350	57	150	0	657270
Animal_372	57	150	0	655890
Animal_374	57	150	0	657180
Animal_379	57	150	0	657240
Animal_401	57	150	0	654150
Animal_425	57	150	0	664920
Animal_433	57	150	0	658705
Animal_435	57	150	0	664022
Animal_443	57	150	0	648515
Animal_445	57	150	0	662100
Animal_466	57	150	0	658741
Animal_478	57	150	0	657534
Animal_479	57	150	0	665760
Animal_500	57	150	0	662100

Animal_508	57	150	0	655500
Animal_515	57	150	0	663818
Animal_519	57	150	0	656190
Animal_541	57	150	0	657545
Animal_603	57	150	0	658890
Animal_617	57	150	0	650266
Animal_635	57	150	0	659430
Animal_638	57	150	0	654180
Animal_644	57	150	0	655680
Animal_653	57	150	0	655650
Animal_670	57	150	0	662328
Animal_672	57	150	0	657289
Animal_677	57	150	0	657259
Animal_693	57	150	0	659428
Animal_700	57	150	0	667318
Animal_718	57	150	0	656160
Animal_726	57	150	0	662250
Animal_738	57	150	0	655830
Animal_749	57	150	0	656400
Animal_759	57	150	0	663330
Animal_763	57	150	0	664860
Animal_770	57	150	0	665749
Animal_775	57	150	0	657180
Animal_776	57	150	0	656400
Animal_791	57	150	0	657120
Animal_804	57	150	0	664303
Animal_822	57	150	0	656987
Animal_823	57	150	0	662738
Animal_842	57	150	0	657240
Animal_881	57	150	0	655558
Animal_889	57	150	0	658769
Animal_901	57	150	0	654545
Animal_911	57	150	0	664110
Animal_918	57	150	0	654600
Animal_920	57	150	0	657240
Animal_922	57	150	0	657660
Animal_938	57	150	0	650700
Animal_956	57	150	0	646320
Animal_959	57	150	0	657030
Animal_964	57	150	0	662490
Animal_965	57	150	0	659340
Animal_991	57	150	0	662388
Environment_112	57	150	0	657150
Environment_138	57	150	0	654138
Environment_147	57	150	0	664950
Environment_148	57	150	0	662536
Environment_165	57	150	0	662940
Environment_170	57	150	0	659218

Environment_175	57	150	0	646496
Environment_186	57	150	0	662490
Environment_187	57	150	0	652950
Environment_212	57	150	0	662585
Environment_221	57	150	0	657240
Environment_243	57	150	0	662901
Environment_247	57	150	0	662290
Environment_258	57	150	0	662440
Environment_260	57	150	0	655901
Environment_272	57	150	0	658711
Environment_320	57	150	0	634912
Environment_327	57	150	0	657270
Environment_352	57	150	0	656222
Environment_366	57	150	0	661440
Environment_381	57	150	0	656670
Environment_408	57	150	0	662460
Environment_418	57	150	0	663090
Environment_458	57	150	0	657270
Environment_468	57	150	0	655882
Environment_481	57	150	0	648151
Environment_483	57	150	0	657011
Environment_502	57	150	0	660328
Environment_507	57	150	0	650370
Environment_512	57	150	0	655631
Environment_526	57	150	0	656940
Environment_528	57	150	0	651840
Environment_534	57	150	0	662483
Environment_544	57	150	0	664933
Environment_566	57	150	0	661260
Environment_569	57	150	0	662167
Environment_573	57	150	0	655533
Environment_574	57	150	0	655380
Environment_576	57	150	0	662510
Environment_591	57	150	0	657727
Environment_595	57	150	0	651990
Environment_597	57	150	0	639399
Environment_609	57	150	0	654174
Environment_619	57	150	0	656670
Environment_637	57	150	0	657000
Environment_648	57	150	0	657270
Environment_651	57	150	0	662019
Environment_665	57	150	0	656790
Environment_695	57	150	0	652726
Environment_706	57	150	0	659628
Environment_714	57	150	0	661993
Environment_728	57	150	0	659520
Environment_730	57	150	0	657240
Environment_735	57	150	0	663487

Environment_750	57	150	0	655650
Environment_757	57	150	0	657210
Environment_759	57	150	0	662446
Environment_774	57	150	0	658728
Environment_779	57	150	0	659674
Environment_788	57	150	0	656040
Environment_790	57	150	0	663270
Environment_796	57	150	0	627636
Environment_814	57	150	0	657030
Environment_818	57	150	0	655950
Environment_838	57	150	0	653130
Environment_839	57	150	0	654554
Environment_848	57	150	0	655901
Environment_862	57	150	0	657136
Environment_875	57	150	0	657210
Environment_883	57	150	0	662250
Environment_888	57	150	0	657180
Environment_902	57	150	0	665730
Environment_905	57	150	0	658737
Environment_907	57	150	0	659542
Environment_908	57	150	0	664890
Environment_917	57	150	0	657360
Environment_924	57	150	0	662220
Environment_932	57	150	0	651060
Environment_943	57	150	0	662405
Environment_955	57	150	0	656040
Environment_963	57	150	0	653932
Environment_966	57	150	0	649215
Environment_998	57	150	0	650730
Food_160	57	150	0	657030
Food_210	57	150	0	659430
Food_215	57	150	0	653299
Food_223	57	150	0	655562
Food_233	57	150	0	659160
Food_237	57	150	0	660570
Food_295	57	150	0	656329
Food_310	57	150	0	657204
Food_314	57	150	0	661166
Food_389	57	150	0	657120
Food_403	57	150	0	665670
Food_452	57	150	0	655622
Food_485	57	150	0	659828
Food_501	57	150	0	662520
Food_525	57	150	0	659854
Food_538	57	150	0	660458
Food_543	57	150	0	657060
Food_585	57	150	0	664939
Food_627	57	150	0	662047

Food_642	57	150	0	655620
Food_659	57	150	0	673704
Food_668	57	150	0	657198
Food_689	57	150	0	655830
Food_702	57	150	0	652500
Food_729	57	150	0	650790
Food_762	57	150	0	656433
Food_781	57	150	0	660792
Food_806	57	150	0	654883
Food_821	57	150	0	663900
Food_843	57	150	0	655566
Food_879	57	150	0	659340
Food_967	57	150	0	662231
Food_978	57	150	0	662190
Food_984	57	150	0	650790
Food_992	57	150	0	656910
Human_107	57	150	0	662047
Human_116	57	150	0	667441
Human_119	57	150	0	662049
Human_126	57	150	0	662061
Human_129	57	150	0	659397
Human_132	57	150	0	652200
Human_174	57	150	0	657270
Human_182	57	150	0	662055
Human_184	57	150	0	662060
Human_204	57	150	0	662056
Human_206	57	150	0	638862
Human_225	57	150	0	663720
Human_264	57	150	0	662042
Human_286	57	150	0	662444
Human_329	57	150	0	662062
Human_337	57	150	0	662055
Human_341	57	150	0	662048
Human_359	57	150	0	662030
Human_364	57	150	0	662520
Human_386	57	150	0	626861
Human_412	57	150	0	662063
Human_430	57	150	0	662048
Human_440	57	150	0	662300
Human_454	57	150	0	662062
Human_462	57	150	0	662057
Human_488	57	150	0	662055
Human_496	57	150	0	669540
Human_518	57	150	0	649936
Human_522	57	150	0	662048
Human_533	57	150	0	651390
Human_546	57	150	0	662066
Human_563	57	150	0	658672

Human_579	57	150	0	662048
Human_581	57	150	0	651210
Human_598	57	150	0	662039
Human_600	57	150	0	662044
Human_605	57	150	0	664260
Human_632	57	150	0	662056
Human_649	57	150	0	662042
Human_655	57	150	0	662062
Human_671	57	150	0	662053
Human_681	57	150	0	657204
Human_694	57	150	0	657180
Human_733	57	150	0	662047
Human_789	57	150	0	662220
Human_793	57	150	0	662052
Human_829	57	150	0	653911
Human_832	57	150	0	656692
Human_835	57	150	0	655147
Human_845	57	150	0	659370
Human_847	57	150	0	657300
Human_854	57	150	0	662053
Human_858	57	150	0	662054
Human_891	57	150	0	664210
Human_928	57	150	0	656100
Human_949	57	150	0	659325
Human_950	57	150	0	652410
Human_974	57	150	0	664050
Human_980	57	150	0	662044
Human_983	57	150	0	649022
Food_sample01	44	35-251	0	1110809
Food_sample02	45	35-251	0	2011134
Food_sample03	44	35-251	0	760316
Food_sample04	46	35-251	0	1690333
Food_sample05	46	35-251	0	1663849
Food_sample06	46	35-251	0	1593886
Food_sample07	46	35-251	0	1735397
Food_sample08	46	35-251	0	2377216
Food_sample09	46	35-251	0	1833908
Food_sample10	46	35-251	0	1992757
Food_sample11	44	35-251	0	1769201
Food_sample12	46	35-251	0	2417483
Food_sample13	44	35-251	0	2831532
Food_sample14	44	35-251	0	1699659

**Tabla Anexo 2.** Parámetros de calidad principales de los ficheros FASTA del Componente2 (parse\_quast.py).

Muestra	Nºcontigs totales	Largest contig <sup>a</sup>	N50 <sup>b</sup>	N75 <sup>c</sup>	L50 <sup>d</sup>	L75 <sup>e</sup>
Animal_113	38	358992	176832	120286	7	13
Animal_127	40	496004	204682	122402	6	11
Animal_140	24	599662	293651	190164	4	7
Animal_143	33	420193	194857	143811	7	11
Animal_154	29	609074	293138	178944	4	8
Animal_173	31	518912	294121	190317	5	8
Animal_189	28	529790	253944	190267	5	8
Animal_209	41	548137	178764	105204	7	13
Animal_217	33	419979	221848	155896	6	11
Animal_222	36	458538	190600	117353	6	11
Animal_234	33	420057	194879	143548	7	11
Animal_236	36	519419	140596	105365	6	13
Animal_241	32	438896	249780	138045	5	10
Animal_251	35	531300	280422	103020	5	11
Animal_254	27	530749	259778	190174	5	8
Animal_267	31	588245	222274	135699	5	10
Animal_268	29	518913	294228	189622	5	8
Animal_275	37	372966	140765	105365	7	13
Animal_279	32	387280	259386	135387	5	10
Animal_285	38	363207	183050	103797	6	13
Animal_287	29	512109	249780	160013	5	9
Animal_289	32	518914	259902	138541	5	9
Animal_290	30	453324	189498	127257	6	11
Animal_317	29	518913	294155	190336	5	8
Animal_325	29	886135	364262	135788	3	8
Animal_331	27	419817	222072	159943	6	10
Animal_350	31	518913	294155	190336	5	8
Animal_372	43	427319	166059	116099	6	11
Animal_374	29	518913	294228	190304	5	8
Animal_379	28	518905	294083	190304	5	8
Animal_401	22	518992	312445	216944	4	7
Animal_425	39	458666	142585	105307	6	12
Animal_433	31	419920	221873	155889	6	11
Animal_435	40	375750	212323	95077	6	12
Animal_443	54	328214	122679	74836	9	18
Animal_445	31	512135	249755	138350	5	9
Animal_466	32	420051	194882	143548	7	11
Animal_478	23	518913	316923	188530	4	7
Animal_479	42	519407	140075	103108	7	14
Animal_500	29	459895	250029	189845	5	9
Animal_508	30	531872	257400	150202	5	8
Animal_515	37	371125	183015	105420	6	13
Animal_519	34	518982	294148	138415	4	8

Animal_541	32	609381	249716	116632	5	10
Animal_603	32	510685	249310	150810	5	9
Animal_617	20	1005899	309982	251870	3	6
Animal_635	24	420661	251081	193396	6	9
Animal_638	25	518991	312440	217284	4	7
Animal_644	32	554021	197272	121813	5	10
Animal_653	29	461507	259778	138493	5	9
Animal_670	32	454269	249754	116413	5	10
Animal_672	22	639240	294068	216941	4	7
Animal_677	30	518913	294073	190305	5	8
Animal_693	24	662746	293390	179255	4	8
Animal_700	38	373024	170153	103772	7	14
Animal_718	25	588306	382163	190237	4	7
Animal_726	31	512676	249544	138045	5	10
Animal_738	40	427282	184312	135158	6	11
Animal_749	32	461732	222044	138063	5	10
Animal_759	45	320080	117214	87653	9	17
Animal_763	38	458566	183689	105295	6	12
Animal_770	40	371032	163604	104007	6	13
Animal_775	32	518913	308894	190305	4	8
Animal_776	32	461708	249647	170186	5	9
Animal_791	30	518911	294074	190304	5	8
Animal_804	40	531590	140077	103128	6	13
Animal_822	31	462128	254102	217226	6	9
Animal_823	31	359117	189612	116357	6	12
Animal_842	35	462258	195547	105465	5	11
Animal_881	26	410840	269516	177526	5	9
Animal_889	33	420109	194895	143560	7	11
Animal_901	34	518926	241882	158724	6	10
Animal_911	38	319985	139131	97062	8	15
Animal_918	25	519005	308408	216955	4	7
Animal_920	30	393571	201422	105491	6	11
Animal_922	30	588243	232491	131102	5	9
Animal_938	33	391187	254264	135324	5	10
Animal_956	22	462370	358831	189512	4	8
Animal_959	39	358867	142294	102877	8	14
Animal_964	27	608883	207679	160283	5	9
Animal_965	35	519425	170357	105327	6	12
Animal_991	28	454200	249672	138327	5	10
Environment_112	29	518912	294098	190317	5	8
Environment_138	90	243999	85412	46204	12	26
Environment_147	39	519323	139159	99044	7	14
Environment_148	31	359133	236632	138407	6	11
Environment_165	36	320082	139354	103213	7	13
Environment_170	26	586243	259710	145720	4	8
Environment_175	130	176319	48778	28434	22	43
Environment_186	28	512893	251116	138327	5	10
Environment_187	34	480623	294015	190149	5	8



Environment_212	31	531946	293573	150601	4	8
Environment_221	31	518914	294064	190293	5	8
Environment_243	34	532032	182395	105345	5	11
Environment_247	31	609301	293316	160043	4	8
Environment_258	31	359117	236465	138326	6	11
Environment_260	26	518980	363730	190175	4	7
Environment_272	33	420201	194889	143560	7	11
Environment_320	429	48838	11712	6729	79	167
Environment_327	33	518913	259919	157082	5	9
Environment_352	27	518741	228767	137586	5	9
Environment_366	32	532013	259304	134810	5	9
Environment_381	30	386102	294067	153517	5	9
Environment_408	29	512717	249753	138327	5	10
Environment_418	36	458569	168289	105366	6	12
Environment_458	31	518913	294215	190304	5	8
Environment_468	31	386418	256571	131575	5	10
Environment_481	49	261000	135331	104691	10	16
Environment_483	30	588310	294280	127669	4	9
Environment_502	29	588323	294243	138270	4	8
Environment_507	25	518776	294021	180235	4	8
Environment_512	28	400608	259946	135553	5	10
Environment_526	33	390030	250187	140260	6	10
Environment_528	23	586274	292306	217953	4	7
Environment_534	31	513054	249750	116377	5	10
Environment_544	40	519254	140595	103172	6	13
Environment_566	25	586268	293989	190203	4	7
Environment_569	32	453330	249763	138046	5	10
Environment_573	27	585059	294177	169103	4	8
Environment_574	94	144578	64905	38800	19	35
Environment_576	37	462235	194129	138353	6	11
Environment_591	79	320063	144290	74875	7	15
Environment_595	21	589812	294131	218031	4	7
Environment_597	25	475405	372177	156074	4	8
Environment_609	24	530645	259864	193618	5	8
Environment_619	49	315387	177563	105216	8	14
Environment_637	30	364405	294097	138541	5	10
Environment_648	31	518914	294243	190305	5	8
Environment_651	30	453295	249542	150230	5	10
Environment_665	30	391343	294089	153517	5	9
Environment_695	34	486813	294017	190165	5	8
Environment_706	25	610034	249452	159942	5	9
Environment_714	29	512021	249558	160005	5	9
Environment_728	32	609394	236581	171807	5	9
Environment_730	29	518906	294076	190300	5	8
Environment_735	36	519265	140620	103114	6	13
Environment_750	25	586628	297667	216852	4	7
Environment_757	32	518913	294096	127748	5	9
Environment_759	30	454373	249753	138327	5	10

Environment_774	24	629784	259769	217081	4	7
Environment_779	23	420380	281637	193360	5	8
Environment_788	61	272228	118826	79199	10	19
Environment_790	35	458645	243479	105423	5	12
Environment_796	22	723440	457912	156034	3	6
Environment_814	29	480982	294212	190266	5	8
Environment_818	29	538428	293651	163448	4	8
Environment_838	23	586203	294068	216965	4	7
Environment_839	23	487182	250086	169970	6	9
Environment_848	34	382195	190237	122377	7	13
Environment_862	33	518911	259926	190254	5	9
Environment_875	30	383722	294098	135611	5	10
Environment_883	31	512007	249763	138046	5	10
Environment_888	29	518912	294194	190306	5	8
Environment_902	32	364405	259768	127740	6	10
Environment_905	33	420325	194872	143585	7	11
Environment_907	23	609266	278268	153353	4	8
Environment_908	37	519408	140620	105365	7	13
Environment_917	34	518907	259907	186208	5	8
Environment_924	30	468721	249790	150265	5	10
Environment_932	31	586359	293932	137938	4	9
Environment_943	32	454503	249753	138327	5	10
Environment_955	32	520089	293651	190212	4	8
Environment_963	25	586273	292298	217120	4	7
Environment_966	84	173401	78199	38783	15	30
Environment_998	28	482594	259509	127694	5	10
Food_160	33	462045	250985	116380	5	11
Food_210	24	439463	250973	159952	5	9
Food_215	116	218428	56816	31409	17	37
Food_223	29	457753	253812	196447	5	9
Food_233	32	518911	259923	140250	5	9
Food_237	40	480731	259886	137736	5	9
Food_295	22	886420	530583	216800	3	5
Food_310	34	401035	251704	156001	5	9
Food_314	23	609387	293520	171817	4	8
Food_389	36	453335	249474	94462	5	12
Food_403	41	498161	140088	105366	7	13
Food_452	27	462473	250080	189207	5	9
Food_485	27	519070	259579	190264	5	8
Food_501	31	629652	293563	190596	4	7
Food_525	23	419986	221549	182654	6	10
Food_538	22	530453	259550	216904	5	8
Food_543	34	452134	178995	105500	6	12
Food_585	40	458731	142555	103801	6	12
Food_627	28	513252	242288	135559	5	10
Food_642	23	639853	457977	222368	3	6
Food_659	48	299117	135549	79365	9	17
Food_668	30	518914	294089	190323	5	8

Food_689	40	358661	189502	135181	6	11
Food_702	28	457905	259950	138177	5	9
Food_729	33	433101	249974	153633	5	9
Food_762	24	518523	259621	216888	5	8
Food_781	22	663541	293670	159953	4	8
Food_806	28	585044	294149	190081	4	8
Food_821	35	518933	251728	139302	5	9
Food_843	27	390252	260897	189401	5	9
Food_879	25	443095	249459	190739	5	9
Food_967	32	453400	249764	138353	5	10
Food_978	29	511997	249788	159986	5	9
Food_984	29	461689	259783	189937	5	8
Food_992	45	358863	140688	74999	9	17
Human_107	28	518613	242288	139705	5	9
Human_116	38	530480	140620	105365	6	12
Human_119	27	518613	242288	139705	5	9
Human_126	27	518613	242288	139705	5	9
Human_129	35	528608	259520	134803	5	8
Human_132	23	590700	294041	190305	4	7
Human_174	32	518914	294097	127749	5	9
Human_182	27	518613	242288	139705	5	9
Human_184	27	518613	242288	139705	5	9
Human_204	27	518613	242288	139420	5	9
Human_206	28	518694	260648	139324	5	9
Human_225	36	371106	166820	105354	7	14
Human_264	27	518613	242288	139705	5	9
Human_286	30	512797	249787	138329	5	10
Human_329	27	518613	242288	139705	5	9
Human_337	27	518613	242288	139705	5	9
Human_341	27	518613	242288	139705	5	9
Human_359	27	518613	242288	139705	5	9
Human_364	29	512935	249745	138327	5	10
Human_386	19	586201	376245	194173	4	6
Human_412	27	518613	242288	139705	5	9
Human_430	27	518613	242288	139705	5	9
Human_440	29	512196	249558	159995	5	9
Human_454	27	518613	242288	139420	5	9
Human_462	27	518613	242288	139705	5	9
Human_488	27	518613	242288	139705	5	9
Human_496	84	292954	138433	112419	9	15
Human_518	58	260616	122117	60453	10	19
Human_522	27	518613	242288	139705	5	9
Human_533	30	586601	293929	137844	4	9
Human_546	27	518613	242288	139420	5	9
Human_563	33	353716	196162	139075	7	12
Human_579	27	518613	242288	139705	5	9
Human_581	32	364273	253845	135509	6	10
Human_598	27	518613	242288	139705	5	9

Human_600	27	518613	242288	139705	5	9
Human_605	36	446399	180235	102846	6	12
Human_632	27	518613	242288	139705	5	9
Human_649	27	518613	242288	139705	5	9
Human_655	27	518613	242288	139705	5	9
Human_671	27	518613	242288	139705	5	9
Human_681	32	518916	294105	190304	5	8
Human_694	31	518914	294081	127749	5	9
Human_733	27	518613	242288	139705	5	9
Human_789	31	512106	249568	138327	5	10
Human_793	27	518613	242288	139705	5	9
Human_829	24	584564	297055	175228	4	8
Human_832	31	395595	186966	139490	6	11
Human_835	24	639569	294014	156061	4	8
Human_845	29	390585	194924	159893	6	10
Human_847	34	518914	294081	127749	5	9
Human_854	27	518613	242288	139705	5	9
Human_858	27	518613	242288	139272	5	9
Human_891	38	541516	222362	139277	5	10
Human_928	30	398044	259980	187156	5	9
Human_949	24	608655	293282	170052	4	8
Human_950	19	591066	294141	259975	4	7
Human_974	26	587442	294141	190148	4	7
Human_980	27	518613	242288	139705	5	9
Human_983	130	156797	48232	32156	24	43

<sup>a</sup> Contig de mayor tamaño. <sup>b</sup> N50 es la longitud para la cual la colección de todos los contigs de esa longitud o más larga cubre al menos la mitad de un ensamblado. <sup>c</sup> N75 se define de manera similar a N50 pero con 75% en lugar de 50%. <sup>d</sup> L50 (L75 <sup>e</sup>) es el número de contigs igual o mayor que N50 (N75).

**Tabla Anexo 3.** Resultados del análisis taxonómico de Kraken y Bracken para muestras del Componente 1. En la tabla aparecen resumidos los parámetros de hit con mayor número de lecturas asignadas (parse\_kraken.py).

Muestra	Fracción total de lecturas	Lecturas asignadas	Nombre	Id Taxómico
Animal_113	0,92341	606620	<i>B. melitensis</i>	29459
Animal_127	0,91839	602274	<i>B. melitensis</i>	29459
Animal_140	0,82686	542307	<i>B. abortus</i>	235
Animal_143	0,91655	603605	<i>B. melitensis</i>	29459
Animal_154	0,93471	613667	<i>B. melitensis</i>	29459
Animal_173	0,72824	478482	<i>B. abortus</i>	235
Animal_189	0,71163	471055	<i>B. abortus</i>	235
Animal_209	0,90428	592907	<i>B. melitensis</i>	29459
Animal_217	0,92398	608463	<i>B. melitensis</i>	29459
Animal_222	0,77433	504916	<i>B. canis</i>	36855
Animal_234	0,93193	613713	<i>B. melitensis</i>	29459
Animal_236	0,93255	618854	<i>B. suis</i>	29461
Animal_241	0,93348	617908	<i>B. melitensis</i>	29459
Animal_251	0,93974	628392	<i>B. suis</i>	29461
Animal_254	0,86102	567922	<i>B. abortus</i>	235
Animal_267	0,94395	620561	<i>B. abortus</i>	235
Animal_268	0,44895	294928	<i>B. melitensis</i>	29459
Animal_275	0,93606	623830	<i>B. suis</i>	29461
Animal_279	0,81929	533217	<i>B. abortus</i>	235
Animal_285	0,93447	620316	<i>B. suis</i>	29461
Animal_287	0,93195	617023	<i>B. melitensis</i>	29459
Animal_289	0,485	318582	<i>B. melitensis</i>	29459
Animal_290	0,92167	605557	<i>B. melitensis</i>	29459
Animal_317	0,87225	573112	<i>B. abortus</i>	235
Animal_325	0,90173	592991	<i>B. abortus</i>	235
Animal_331	0,93159	614327	<i>B. melitensis</i>	29459
Animal_350	0,50845	334085	<i>B. abortus</i>	235
Animal_372	0,91002	596699	<i>B. melitensis</i>	29459
Animal_374	0,44857	294701	<i>B. melitensis</i>	29459
Animal_379	0,6697	440010	<i>B. abortus</i>	235
Animal_401	0,87301	570903	<i>B. abortus</i>	235
Animal_425	0,95404	634152	<i>B. suis</i>	29461
Animal_433	0,93	612415	<i>B. melitensis</i>	29459
Animal_435	0,95022	630803	<i>B. suis</i>	29461
Animal_443	0,84871	550238	<i>B. abortus</i>	235
Animal_445	0,93145	616508	<i>B. melitensis</i>	29459
Animal_466	0,91974	605683	<i>B. melitensis</i>	29459
Animal_478	0,42216	277500	<i>B. abortus</i>	235
Animal_479	0,94073	626123	<i>B. suis</i>	29461
Animal_500	0,9176	607355	<i>B. melitensis</i>	29459
Animal_508	0,83715	548571	<i>B. canis</i>	36855

Animal_515	0,93055	617512	<i>B. suis</i>	29461
Animal_519	0,91939	603118	<i>B. abortus</i>	235
Animal_541	0,90992	598134	<i>B. melitensis</i>	29459
Animal_603	0,95682	630233	<i>B. melitensis</i>	29459
Animal_617	0,50319	327107	<i>B. abortus</i>	235
Animal_635	0,91216	601323	<i>B. melitensis</i>	29459
Animal_638	0,86365	564793	<i>B. abortus</i>	235
Animal_644	0,89987	589838	<i>B. abortus</i>	235
Animal_653	0,88527	580268	<i>B. abortus</i>	235
Animal_670	0,91327	604699	<i>B. melitensis</i>	29459
Animal_672	0,84516	555311	<i>B. abortus</i>	235
Animal_677	0,7252	476483	<i>B. abortus</i>	235
Animal_693	0,91464	602920	<i>B. melitensis</i>	29459
Animal_700	0,93202	621753	<i>B. suis</i>	29461
Animal_718	0,90899	596243	<i>B. abortus</i>	235
Animal_726	0,93548	619357	<i>B. melitensis</i>	29459
Animal_738	0,91742	601471	<i>B. melitensis</i>	29459
Animal_749	0,93221	611690	<i>B. melitensis</i>	29459
Animal_759	0,95018	630093	<i>B. suis</i>	29461
Animal_763	0,95038	631684	<i>B. suis</i>	29461
Animal_770	0,93253	620627	<i>B. suis</i>	29461
Animal_775	0,74963	492496	<i>B. abortus</i>	235
Animal_776	0,91825	602549	<i>B. melitensis</i>	29459
Animal_791	0,76689	503771	<i>B. abortus</i>	235
Animal_804	0,94695	628866	<i>B. suis</i>	29461
Animal_822	0,8647	567937	<i>B. abortus</i>	235
Animal_823	0,92073	610040	<i>B. melitensis</i>	29459
Animal_842	0,93518	614464	<i>B. melitensis</i>	29459
Animal_881	0,92679	607364	<i>B. abortus</i>	235
Animal_889	0,9185	604900	<i>B. melitensis</i>	29459
Animal_901	0,68051	445269	<i>B. abortus</i>	235
Animal_911	0,93809	622765	<i>B. suis</i>	29461
Animal_918	0,84014	549778	<i>B. abortus</i>	235
Animal_920	0,92369	606907	<i>B. melitensis</i>	29459
Animal_922	0,93267	613194	<i>B. abortus</i>	235
Animal_938	0,82365	535768	<i>B. abortus</i>	235
Animal_956	0,92536	597927	<i>B. melitensis</i>	29459
Animal_959	0,93706	615508	<i>B. melitensis</i>	29459
Animal_964	0,93666	620301	<i>B. melitensis</i>	29459
Animal_965	0,94561	623281	<i>B. suis</i>	29461
Animal_991	0,91942	608824	<i>B. melitensis</i>	29459
Environment_112	0,60636	398347	<i>B. abortus</i>	235
Environment_138	0,83641	546950	<i>B. canis</i>	36855
Environment_147	0,95734	636388	<i>B. suis</i>	29461
Environment_148	0,91162	603790	<i>B. melitensis</i>	29459
Environment_165	0,94297	624954	<i>B. suis</i>	29461
Environment_170	0,86351	569072	<i>B. abortus</i>	235
Environment_175	0,84154	543878	<i>B. abortus</i>	235

Environment_186	0,91434	605565	<i>B. melitensis</i>	29459
Environment_187	0,72358	472305	<i>B. abortus</i>	235
Environment_212	0,79575	527077	<i>B. canis</i>	36855
Environment_221	0,87102	572281	<i>B. abortus</i>	235
Environment_243	0,96314	638259	<i>B. suis</i>	29461
Environment_247	0,93509	619132	<i>B. melitensis</i>	29459
Environment_258	0,91051	602970	<i>B. melitensis</i>	29459
Environment_260	0,8473	555586	<i>B. abortus</i>	235
Environment_272	0,93275	614251	<i>B. melitensis</i>	29459
Environment_320	0,92735	588602	<i>B. melitensis</i>	29459
Environment_327	0,52716	346387	<i>B. melitensis</i>	29459
Environment_352	0,8664	568340	<i>B. abortus</i>	235
Environment_366	0,8491	561450	<i>B. canis</i>	36855
Environment_381	0,46602	305926	<i>B. abortus</i>	235
Environment_408	0,9243	612126	<i>B. melitensis</i>	29459
Environment_418	0,93758	621513	<i>B. suis</i>	29461
Environment_458	0,41961	275707	<i>B. melitensis</i>	29459
Environment_468	0,88813	582315	<i>B. abortus</i>	235
Environment_481	0,87723	568407	<i>B. abortus</i>	235
Environment_483	0,90894	596983	<i>B. abortus</i>	235
Environment_502	0,89381	590041	<i>B. abortus</i>	235
Environment_507	0,70267	456844	<i>B. abortus</i>	235
Environment_512	0,83987	550478	<i>B. abortus</i>	235
Environment_526	0,48963	321541	<i>B. melitensis</i>	29459
Environment_528	0,82717	539017	<i>B. abortus</i>	235
Environment_534	0,91714	607410	<i>B. melitensis</i>	29459
Environment_544	0,95367	633920	<i>B. suis</i>	29461
Environment_566	0,86048	568841	<i>B. abortus</i>	235
Environment_569	0,93093	616258	<i>B. melitensis</i>	29459
Environment_573	0,89623	587279	<i>B. abortus</i>	235
Environment_574	0,93949	615525	<i>B. melitensis</i>	29459
Environment_576	0,96526	639312	<i>B. melitensis</i>	29459
Environment_591	0,83565	549454	<i>B. abortus</i>	235
Environment_595	0,88025	573752	<i>B. abortus</i>	235
Environment_597	0,82351	526379	<i>B. abortus</i>	235
Environment_609	0,48168	315013	<i>B. melitensis</i>	29459
Environment_619	0,78415	514751	<i>B. melitensis</i>	29459
Environment_637	0,43974	288825	<i>B. melitensis</i>	29459
Environment_648	0,4236	278341	<i>B. abortus</i>	235
Environment_651	0,92903	614864	<i>B. melitensis</i>	29459
Environment_665	0,46011	302091	<i>B. melitensis</i>	29459
Environment_695	0,71947	469461	<i>B. abortus</i>	235
Environment_706	0,91498	603375	<i>B. melitensis</i>	29459
Environment_714	0,93748	620419	<i>B. melitensis</i>	29459
Environment_728	0,92386	609124	<i>B. melitensis</i>	29459
Environment_730	0,47861	314457	<i>B. melitensis</i>	29459
Environment_735	0,95518	633560	<i>B. suis</i>	29461
Environment_750	0,85422	559907	<i>B. abortus</i>	235

Environment_757	0,53618	352258	<i>B. suis</i>	29461
Environment_759	0,91786	607834	<i>B. melitensis</i>	29459
Environment_774	0,82936	546136	<i>B. abortus</i>	235
Environment_779	0,91535	603653	<i>B. melitensis</i>	29459
Environment_788	0,93455	612921	<i>B. melitensis</i>	29459
Environment_790	0,94856	628975	<i>B. suis</i>	29461
Environment_796	0,8737	548194	<i>B. abortus</i>	235
Environment_814	0,47164	309773	<i>B. melitensis</i>	29459
Environment_818	0,91337	598927	<i>B. abortus</i>	235
Environment_838	0,84814	553782	<i>B. abortus</i>	235
Environment_839	0,88172	576955	<i>B. abortus</i>	235
Environment_848	0,92882	609018	<i>B. abortus</i>	235
Environment_862	0,88182	579316	<i>B. abortus</i>	235
Environment_875	0,47993	315310	<i>B. melitensis</i>	29459
Environment_883	0,93114	616464	<i>B. melitensis</i>	29459
Environment_888	0,42963	282247	<i>B. melitensis</i>	29459
Environment_902	0,50951	339081	<i>B. melitensis</i>	29459
Environment_905	0,91904	605223	<i>B. melitensis</i>	29459
Environment_907	0,93027	613386	<i>B. melitensis</i>	29459
Environment_908	0,96055	638436	<i>B. suis</i>	29461
Environment_917	0,53607	352279	<i>B. melitensis</i>	29459
Environment_924	0,95848	634529	<i>B. melitensis</i>	29459
Environment_932	0,81616	531186	<i>B. abortus</i>	235
Environment_943	0,91662	606990	<i>B. melitensis</i>	29459
Environment_955	0,83135	545234	<i>B. abortus</i>	235
Environment_963	0,8548	558817	<i>B. abortus</i>	235
Environment_966	0,81922	531691	<i>B. abortus</i>	235
Environment_998	0,73944	481010	<i>B. abortus</i>	235
Food_160	0,92549	607891	<i>B. melitensis</i>	29459
Food_210	0,92866	612213	<i>B. melitensis</i>	29459
Food_215	0,88522	578119	<i>B. suis</i>	29461
Food_223	0,6537	428411	<i>B. abortus</i>	235
Food_233	0,72375	476923	<i>B. abortus</i>	235
Food_237	0,53988	356485	<i>B. melitensis</i>	29459
Food_295	0,77568	508931	<i>B. abortus</i>	235
Food_310	0,58185	382273	<i>B. melitensis</i>	29459
Food_314	0,92708	612768	<i>B. melitensis</i>	29459
Food_389	0,92613	608409	<i>B. melitensis</i>	29459
Food_403	0,9549	635488	<i>B. suis</i>	29461
Food_452	0,76548	501727	<i>B. abortus</i>	235
Food_485	0,9253	610352	<i>B. abortus</i>	235
Food_501	0,79787	528435	<i>B. canis</i>	36855
Food_525	0,93145	614465	<i>B. melitensis</i>	29459
Food_538	0,76125	502626	<i>B. abortus</i>	235
Food_543	0,92649	608556	<i>B. melitensis</i>	29459
Food_585	0,96331	640356	<i>B. suis</i>	29461
Food_627	0,9335	617810	<i>B. melitensis</i>	29459
Food_642	0,88191	578003	<i>B. abortus</i>	235



Food_659	0,85147	573460	<i>B. abortus</i>	235
Food_668	0,78856	518082	<i>B. abortus</i>	235
Food_689	0,92169	604303	<i>B. melitensis</i>	29459
Food_702	0,81634	532513	<i>B. abortus</i>	235
Food_729	0,82425	536239	<i>B. abortus</i>	235
Food_762	0,79429	521255	<i>B. abortus</i>	235
Food_781	0,90731	599328	<i>B. melitensis</i>	29459
Food_806	0,7781	509401	<i>B. abortus</i>	235
Food_821	0,88886	589908	<i>B. suis</i>	29461
Food_843	0,8308	544482	<i>B. abortus</i>	235
Food_879	0,92429	609243	<i>B. melitensis</i>	29459
Food_967	0,92772	614180	<i>B. melitensis</i>	29459
Food_978	0,92132	609911	<i>B. melitensis</i>	29459
Food_984	0,82899	539313	<i>B. abortus</i>	235
Food_992	0,9184	603112	<i>B. melitensis</i>	29459
Human_107	0,9169	606832	<i>B. melitensis</i>	29459
Human_116	0,95574	637705	<i>B. suis</i>	29461
Human_119	0,93208	616876	<i>B. melitensis</i>	29459
Human_126	0,92288	610828	<i>B. melitensis</i>	29459
Human_129	0,88112	580821	<i>B. suis</i>	29461
Human_132	0,84401	550287	<i>B. abortus</i>	235
Human_174	0,41845	274949	<i>B. melitensis</i>	29459
Human_182	0,92723	613696	<i>B. melitensis</i>	29459
Human_184	0,931	616171	<i>B. melitensis</i>	29459
Human_204	0,92566	612630	<i>B. melitensis</i>	29459
Human_206	0,85657	547060	<i>B. abortus</i>	235
Human_225	0,93552	620737	<i>B. suis</i>	29461
Human_264	0,91084	602835	<i>B. melitensis</i>	29459
Human_286	0,91707	607329	<i>B. melitensis</i>	29459
Human_329	0,92818	614320	<i>B. melitensis</i>	29459
Human_337	0,93132	616401	<i>B. melitensis</i>	29459
Human_341	0,92132	609766	<i>B. melitensis</i>	29459
Human_359	0,92409	611581	<i>B. melitensis</i>	29459
Human_364	0,92335	611573	<i>B. melitensis</i>	29459
Human_386	0,87471	548154	<i>B. abortus</i>	235
Human_412	0,92718	613654	<i>B. melitensis</i>	29459
Human_430	0,92817	614288	<i>B. melitensis</i>	29459
Human_440	0,92056	609515	<i>B. melitensis</i>	29459
Human_454	0,92466	611986	<i>B. melitensis</i>	29459
Human_462	0,92303	610910	<i>B. melitensis</i>	29459
Human_488	0,92794	614166	<i>B. melitensis</i>	29459
Human_496	0,93002	622415	<i>B. melitensis</i>	29459
Human_518	0,92702	602297	<i>B. abortus</i>	235
Human_522	0,94061	622505	<i>B. melitensis</i>	29459
Human_533	0,81311	529495	<i>B. abortus</i>	235
Human_546	0,9257	612674	<i>B. melitensis</i>	29459
Human_563	0,92488	608994	<i>B. melitensis</i>	29459
Human_579	0,93201	616828	<i>B. melitensis</i>	29459

Human_581	0,83656	544612	<i>B. abortus</i>	235
Human_598	0,92561	612563	<i>B. melitensis</i>	29459
Human_600	0,93076	615965	<i>B. melitensis</i>	29459
Human_605	0,84511	561194	<i>B. abortus</i>	235
Human_632	0,9261	612909	<i>B. melitensis</i>	29459
Human_649	0,92363	611290	<i>B. melitensis</i>	29459
Human_655	0,93105	616231	<i>B. melitensis</i>	29459
Human_671	0,93306	617547	<i>B. melitensis</i>	29459
Human_681	0,42829	281381	<i>B. melitensis</i>	29459
Human_694	0,44459	292091	<i>B. abortus</i>	235
Human_733	0,92971	615304	<i>B. melitensis</i>	29459
Human_789	0,92427	611875	<i>B. melitensis</i>	29459
Human_793	0,93103	616200	<i>B. melitensis</i>	29459
Human_829	0,51801	338598	<i>B. melitensis</i>	29459
Human_832	0,37474	245990	<i>B. pinnipedialis</i>	120576
Human_835	0,86906	569199	<i>B. abortus</i>	235
Human_845	0,93446	615992	<i>B. melitensis</i>	29459
Human_847	0,54142	355770	<i>B. melitensis</i>	29459
Human_854	0,91924	608381	<i>B. melitensis</i>	29459
Human_858	0,92891	614807	<i>B. melitensis</i>	29459
Human_891	0,76549	508281	<i>B. abortus</i>	235
Human_928	0,80456	527704	<i>B. abortus</i>	235
Human_949	0,92963	612762	<i>B. melitensis</i>	29459
Human_950	0,8451	551167	<i>B. abortus</i>	235
Human_974	0,80732	535938	<i>B. abortus</i>	235
Human_980	0,93276	617337	<i>B. melitensis</i>	29459
Human_983	0,88819	576289	<i>B. suis</i>	29461

**Tabla Anexo 4.** Resultados del análisis taxonómico de Kraken y Bracken para muestras del Componente 2. En la tabla aparecen resumidos los parámetros del best y second hit con mayor número de lecturas asignadas (parse\_kraken.py).

Muestra	Nombre best_hit	Muestras asignadas best_hit	Fracción total de lecturas best_hit	Nombre second_hit	Muestras asignadas second_hit	Fracción total de lecturas second_hit	Total_hits_06_kraken
Food_sample01	<i>Homo sapiens</i>	49639	0,35103	Actinoalloteichus sp. AHMU CJ021	46453	0,32849	979
Food_sample02	<i>B. abortus</i>	162932	0,29849	<i>B. melitensis</i>	158956	0,29121	1455
Food_sample03	<i>Actinoalloteichus</i> sp. AHMU CJ021	34091	0,31352	<i>Homo sapiens</i>	33323	0,30646	713
Food_sample04	<i>B. melitensis</i>	348692	0,62276	<i>Homo sapiens</i>	63441	0,11331	1145
Food_sample05	<i>B. suis</i>	359998	0,61643	<i>Homo sapiens</i>	59015	0,10105	1125
Food_sample06	<i>B. abortus</i>	165406	0,29142	<i>B. melitensis</i>	151704	0,26728	1064
Food_sample07	<i>B. canis</i>	304926	0,51249	Actinoalloteichus sp. AHMU CJ021	65649	0,11034	1064
Food_sample08	<i>B. suis</i>	352004	0,52274	Actinoalloteichus sp. AHMU CJ021	98812	0,14674	1435
Food_sample09	<i>B. abortus</i>	360135	0,62245	<i>Homo sapiens</i>	68872	0,11904	1147
Food_sample10	<i>B. melitensis</i>	379451	0,63621	<i>Homo sapiens</i>	77429	0,12982	1262
Food_sample11	<i>Homo sapiens</i>	77659	0,35882	Actinoalloteichus sp. AHMU CJ021	71611	0,33087	1441
Food_sample12	<i>B. melitensis</i>	364864	0,55097	Actinoalloteichus sp. AHMU CJ021	94629	0,1429	1513
Food_sample13	<i>B. melitensis</i>	356277	0,45162	<i>Lactococcus lactis</i>	118757	0,15054	1729
Food_sample14	<i>Homo sapiens</i>	75812	0,3595	Actinoalloteichus sp. AHMU CJ021	70715	0,33533	1311

**Tabla Anexo 5.** Resultados del análisis de tipificación multilocus de secuencias.

Muestra	ST	aroA	cobQ	dna <sub>K</sub>	gap	glk	gyrB	int_hyp	omp2 <sub>5</sub>	trpE
Animal_113	39	2	3	2	3	24	1	2	8	5
Animal_127	8	2	3	2	3	3	1	2	8	5
Animal_140	2	1	1	2	2	2	1	1	1	3
Animal_143	8	2	3	2	3	3	1	2	8	5
Animal_154	8	2	3	2	3	3	1	2	8	5
Animal_173	1	1	1	2	2	1	1	1	1	3
Animal_189	1	1	1	2	2	1	1	1	1	3
Animal_209	8	2	3	2	3	3	1	2	8	5
Animal_217	8	2	3	2	3	3	1	2	8	5
Animal_222	21	6	5	1	1	4	5	4	5	3
Animal_234	8	2	3	2	3	3	1	2	8	5
Animal_236	15	2	5	1	1	7	3	3	2	3
Animal_241	8	2	3	2	3	3	1	2	8	5
Animal_251	16	2	5	1	4	7	3	3	2	3
Animal_254	2	1	1	2	2	2	1	1	1	3
Animal_267	6	7	7	7	5	10	6	1	1	3
Animal_268	1	1	1	2	2	1	1	1	1	3
Animal_275	45	10	5	1	1	7	3	3	2	3
Animal_279	2	1	1	2	2	2	1	1	1	3
Animal_285	15	2	5	1	1	7	3	3	2	3
Animal_287	8	2	3	2	3	3	1	2	8	5
Animal_289	1	1	1	2	2	1	1	1	1	3
Animal_290	39	2	3	2	3	24	1	2	8	5
Animal_317	1	1	1	2	2	1	1	1	1	3
Animal_325	32	7	7	7	5	4	1	1	1	3
Animal_331	8	2	3	2	3	3	1	2	8	5
Animal_350	1	1	1	2	2	1	1	1	1	3
Animal_372	8	2	3	2	3	3	1	2	8	5
Animal_374	1	1	1	2	2	1	1	1	1	3
Animal_379	1	1	1	2	2	1	1	1	1	3
Animal_401	2	1	1	2	2	2	1	1	1	3
Animal_425	15	2	5	1	1	7	3	3	2	3
Animal_433	8	2	3	2	3	3	1	2	8	5
Animal_435	15	2	5	1	1	7	3	3	2	3
Animal_443	4	1	1	2	2	2	2	1	1	3
Animal_445	8	2	3	2	3	3	1	2	8	5
Animal_466	8	2	3	2	3	3	1	2	8	5
Animal_478	28	1	12	2	2	1	1	1	1	3
Animal_479	15	2	5	1	1	7	3	3	2	3
Animal_500	8	2	3	2	3	3	1	2	8	5
Animal_508	21	6	5	1	1	4	5	4	5	3
Animal_515	46	2	5	1	4	7	3	3	16	8
Animal_519	4	1	1	2	2	2	2	1	1	3

Animal_541	39	2	3	2	3	24	1	2	8	5
Animal_603	8	2	3	2	3	3	1	2	8	5
Animal_617	Nove l	1	1	2	2	37	1	1	1	3
Animal_635	8	2	3	2	3	3	1	2	8	5
Animal_638	2	1	1	2	2	2	1	1	1	3
Animal_644	33	8	7	7	5	4	1	1	12	3
Animal_653	2	1	1	2	2	2	1	1	1	3
Animal_670	8	2	3	2	3	3	1	2	8	5
Animal_672	2	1	1	2	2	2	1	1	1	3
Animal_677	1	1	1	2	2	1	1	1	1	3
Animal_693	39	2	3	2	3	24	1	2	8	5
Animal_700	16	2	5	1	4	7	3	3	2	3
Animal_718	34	8	7	7	5	4	1	1	1	3
Animal_726	71	2	3	2	3	36	1	2	8	5
Animal_738	8	2	3	2	3	3	1	2	8	5
Animal_749	8	2	3	2	3	3	1	2	8	5
Animal_759	15	2	5	1	1	7	3	3	2	3
Animal_763	15	2	5	1	1	7	3	3	2	3
Animal_770	16	2	5	1	4	7	3	3	2	3
Animal_775	1	1	1	2	2	1	1	1	1	3
Animal_776	8	2	3	2	3	3	1	2	8	5
Animal_791	1	1	1	2	2	1	1	1	1	3
Animal_804	44	2	5	1	1	7	3	3	2	9
Animal_822	4	1	1	2	2	2	2	1	1	3
Animal_823	8	2	3	2	3	3	1	2	8	5
Animal_842	8	2	3	2	3	3	1	2	8	5
Animal_881	6	7	7	7	5	10	6	1	1	3
Animal_889	8	2	3	2	3	3	1	2	8	5
Animal_901	1	1	1	2	2	1	1	1	1	3
Animal_911	16	2	5	1	4	7	3	3	2	3
Animal_918	2	1	1	2	2	2	1	1	1	3
Animal_920	39	2	3	2	3	24	1	2	8	5
Animal_922	6	7	7	7	5	10	6	1	1	3
Animal_938	2	1	1	2	2	2	1	1	1	3
Animal_956	42	2	3	2	3	3	1	2	8	12
Animal_959	39	2	3	2	3	24	1	2	8	5
Animal_964	8	2	3	2	3	3	1	2	8	5
Animal_965	15	2	5	1	1	7	3	3	2	3
Animal_991	8	2	3	2	3	3	1	2	8	5
Environment_112	1	1	1	2	2	1	1	1	1	3
Environment_138	20	6	5	1	1	4	5	4	6	3
Environment_147	15	2	5	1	1	7	3	3	2	3
Environment_148	8	2	3	2	3	3	1	2	8	5
Environment_165	15	2	5	1	1	7	3	3	2	3
Environment_170	2	1	1	2	2	2	1	1	1	3
Environment_175	1	1	1	2	2	1	1	1	1	3

Environment_186	8	2	3	2	3	3	1	2	8	5
Environment_187	1	1	1	2	2	1	1	1	1	3
Environment_212	20	6	5	1	1	4	5	4	6	3
Environment_221	1	1	1	2	2	1	1	1	1	3
Environment_243	15	2	5	1	1	7	3	3	2	3
Environment_247	8	2	3	2	3	3	1	2	8	5
Environment_258	Nove l	2	3	2	3	35	1	2	8	5
Environment_260	2	1	1	2	2	2	1	1	1	3
Environment_272	8	2	3	2	3	3	1	2	8	5
Environment_320	8	2	3	2	3	3	1	2	8	5
Environment_327	1	1	1	2	2	1	1	1	1	3
Environment_352	36	7	9	2	7	4	1	1	1	3
Environment_366	20	6	5	1	1	4	5	4	6	3
Environment_381	1	1	1	2	2	1	1	1	1	3
Environment_408	8	2	3	2	3	3	1	2	8	5
Environment_418	15	2	5	1	1	7	3	3	2	3
Environment_458	1	1	1	2	2	1	1	1	1	3
Environment_468	34	8	7	7	5	4	1	1	1	3
Environment_481	2	1	1	2	2	2	1	1	1	3
Environment_483	34	8	7	7	5	4	1	1	1	3
Environment_502	34	8	7	7	5	4	1	1	1	3
Environment_507	1	1	1	2	2	1	1	1	1	3
Environment_512	2	1	1	2	2	2	1	1	1	3
Environment_526	1	1	1	2	2	1	1	1	1	3
Environment_528	2	1	1	2	2	2	1	1	1	3
Environment_534	8	2	3	2	3	3	1	2	8	5
Environment_544	15	2	5	1	1	7	3	3	2	3
Environment_566	2	1	1	2	2	2	1	1	1	3
Environment_569	8	2	3	2	3	3	1	2	8	5
Environment_573	36	7	9	2	7	4	1	1	1	3
Environment_574	8	2	3	2	3	3	1	2	8	5
Environment_576	8	2	3	2	3	3	1	2	8	5
Environment_591	2	1	1	2	2	2	1	1	1	3
Environment_595	2	1	1	2	2	2	1	1	1	3
Environment_597	1	1	1	2	2	1	1	1	1	3
Environment_609	1	1	1	2	2	1	1	1	1	3
Environment_619	8	2	3	2	3	3	1	2	8	5
Environment_637	1	1	1	2	2	1	1	1	1	3
Environment_648	1	1	1	2	2	1	1	1	1	3
Environment_651	8	2	3	2	3	3	1	2	8	5
Environment_665	1	1	1	2	2	1	1	1	1	3
Environment_695	1	1	1	2	2	1	1	1	1	3
Environment_706	8	2	3	2	3	3	1	2	8	5
Environment_714	8	2	3	2	3	3	1	2	8	5
Environment_728	39	2	3	2	3	24	1	2	8	5
Environment_730	1	1	1	2	2	1	1	1	1	3

Environment_735	15	2	5	1	1	7	3	3	2	3
Environment_750	2	1	1	2	2	2	1	1	1	3
Environment_757	1	1	1	2	2	1	1	1	1	3
Environment_759	8	2	3	2	3	3	1	2	8	5
Environment_774	2	1	1	2	2	2	1	1	1	3
Environment_779	8	2	3	2	3	3	1	2	8	5
Environment_788	8	2	3	2	3	3	1	2	8	5
Environment_790	15	2	5	1	1	7	3	3	2	3
Environment_796	2	1	1	2	2	2	1	1	1	3
Environment_814	1	1	1	2	2	1	1	1	1	3
Environment_818	2	1	1	2	2	2	1	1	1	3
Environment_838	2	1	1	2	2	2	1	1	1	3
Environment_839	2	1	1	2	2	2	1	1	1	3
Environment_848	34	8	7	7	5	4	1	1	1	3
Environment_862	1	1	1	2	2	1	1	1	1	3
Environment_875	1	1	1	2	2	1	1	1	1	3
Environment_883	8	2	3	2	3	3	1	2	8	5
Environment_888	1	1	1	2	2	1	1	1	1	3
Environment_902	1	1	1	2	2	1	1	1	1	3
Environment_905	8	2	3	2	3	3	1	2	8	5
Environment_907	8	2	3	2	3	3	1	2	8	5
Environment_908	15	2	5	1	1	7	3	3	2	3
Environment_917	1	1	1	2	2	1	1	1	1	3
Environment_924	8	2	3	2	3	3	1	2	8	5
Environment_932	Nove l	1	1	2	22	2	1	1	1	3
Environment_943	8	2	3	2	3	3	1	2	8	5
Environment_955	2	1	1	2	2	2	1	1	1	3
Environment_963	2	1	1	2	2	2	1	1	1	3
Environment_966	2	1	1	2	2	2	1	1	1	3
Environment_998	1	1	1	2	2	1	1	1	1	3
Food_160	8	2	3	2	3	3	1	2	8	5
Food_210	8	2	3	2	3	3	1	2	8	5
Food_215	19	2	5	6	1	4	1	1	2	3
Food_223	1	1	1	2	2	1	1	1	1	3
Food_233	1	1	1	2	2	1	1	1	1	3
Food_237	1	1	1	2	2	1	1	1	1	3
Food_295	2	1	1	2	2	2	1	1	1	3
Food_310	1	1	1	2	2	1	1	1	1	3
Food_314	40	2	3	2	3	25	1	2	8	5
Food_389	39	2	3	2	3	24	1	2	8	5
Food_403	15	2	5	1	1	7	3	3	2	3
Food_452	2	1	1	2	2	2	1	1	1	3
Food_485	6	7	7	7	5	10	6	1	1	3
Food_501	20	6	5	1	1	4	5	4	6	3
Food_525	8	2	3	2	3	3	1	2	8	5
Food_538	2	1	1	2	2	2	1	1	1	3

Food_543	39	2	3	2	3	24	1	2	8	5
Food_585	15	2	5	1	1	7	3	3	2	3
Food_627	11	2	3	2	3	3	1	2	10	5
Food_642	35	7	9	2	7	4	1	1	20	3
Food_659	3	1	1	2	6	2	1	1	1	3
Food_668	1	1	1	2	2	1	1	1	1	3
Food_689	8	2	3	2	3	3	1	2	8	5
Food_702	2	1	1	2	2	2	1	1	1	3
Food_729	2	1	1	2	2	2	1	1	1	3
Food_762	2	1	1	2	2	2	1	1	1	3
Food_781	8	2	3	2	3	3	1	2	8	5
Food_806	31	1	1	2	2	2	1	1	17	3
Food_821	19	2	5	6	1	4	1	1	2	3
Food_843	2	1	1	2	2	2	1	1	1	3
Food_879	8	2	3	2	3	3	1	2	8	5
Food_967	8	2	3	2	3	3	1	2	8	5
Food_978	8	2	3	2	3	3	1	2	8	5
Food_984	2	1	1	2	2	2	1	1	1	3
Food_992	39	2	3	2	3	24	1	2	8	5
Human_107	11	2	3	2	3	3	1	2	10	5
Human_116	15	2	5	1	1	7	3	3	2	3
Human_119	11	2	3	2	3	3	1	2	10	5
Human_126	11	2	3	2	3	3	1	2	10	5
Human_129	17	6	5	1	1	4	5	4	2	3
Human_132	30	1	1	2	2	2	9	1	1	3
Human_174	1	1	1	2	2	1	1	1	1	3
Human_182	11	2	3	2	3	3	1	2	10	5
Human_184	11	2	3	2	3	3	1	2	10	5
Human_204	11	2	3	2	3	3	1	2	10	5
Human_206	1	1	1	2	2	1	1	1	1	3
Human_225	16	2	5	1	4	7	3	3	2	3
Human_264	11	2	3	2	3	3	1	2	10	5
Human_286	8	2	3	2	3	3	1	2	8	5
Human_329	11	2	3	2	3	3	1	2	10	5
Human_337	11	2	3	2	3	3	1	2	10	5
Human_341	11	2	3	2	3	3	1	2	10	5
Human_359	11	2	3	2	3	3	1	2	10	5
Human_364	8	2	3	2	3	3	1	2	8	5
Human_386	2	1	1	2	2	2	1	1	1	3
Human_412	11	2	3	2	3	3	1	2	10	5
Human_430	11	2	3	2	3	3	1	2	10	5
Human_440	8	2	3	2	3	3	1	2	8	5
Human_454	11	2	3	2	3	3	1	2	10	5
Human_462	11	2	3	2	3	3	1	2	10	5
Human_488	11	2	3	2	3	3	1	2	10	5
Human_496	8	2	3	2	3	3	1	2	8	5



Human_518	6	7	7	7	5	10	6	1	1	3
Human_522	11	2	3	2	3	3	1	2	10	5
Human_533	2	1	1	2	2	2	1	1	1	3
Human_546	11	2	3	2	3	3	1	2	10	5
Human_563	8	2	3	2	3	3	1	2	8	5
Human_579	11	2	3	2	3	3	1	2	10	5
Human_581	2	1	1	2	2	2	1	1	1	3
Human_598	11	2	3	2	3	3	1	2	10	5
Human_600	11	2	3	2	3	3	1	2	10	5
Human_605	2	1	1	2	2	2	1	1	1	3
Human_632	11	2	3	2	3	3	1	2	10	5
Human_649	11	2	3	2	3	3	1	2	10	5
Human_655	11	2	3	2	3	3	1	2	10	5
Human_671	11	2	3	2	3	3	1	2	10	5
Human_681	1	1	1	2	2	1	1	1	1	3
Human_694	1	1	1	2	2	1	1	1	1	3
Human_733	11	2	3	2	3	3	1	2	10	5
Human_789	8	2	3	2	3	3	1	2	8	5
Human_793	11	2	3	2	3	3	1	2	10	5
Human_829	37	2	9	8	2	19	8	1	14	3
Human_832	38	2	9	8	2	16	8	1	13	3
Human_835	2	1	1	2	2	2	1	1	1	3
Human_845	41	2	3	2	3	26	1	2	8	5
Human_847	1	1	1	2	2	1	1	1	1	3
Human_854	11	2	3	2	3	3	1	2	10	5
Human_858	11	2	3	2	3	3	1	2	10	5
Human_891	2	1	1	2	2	2	1	1	1	3
Human_928	2	1	1	2	2	2	1	1	1	3
Human_949	8	2	3	2	3	3	1	2	8	5
Human_950	2	1	1	2	2	2	1	1	1	3
Human_974	2	1	1	2	2	2	1	1	1	3
Human_980	11	2	3	2	3	3	1	2	10	5
Human_983	17	6	5	1	1	4	5	4	2	3

